

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Lifeline for livers

System for preserving
transplant organs ditches
ice in favour of perfused
nutrients at body
temperature

PAGES 40 & 50

PSYCHOLOGY

CONFLICT RESOLUTION

Can science help Colombia
escape its years of violence?

PAGE 19

ASTROPHYSICS

SECRETS OF A GAS GIANT

First detection of helium in
an exoplanet's atmosphere

PAGES 35 & 68

PALAEONTOLOGY

LEARNING TO FLY

Skull illuminates transition
from dinosaurs to birds

PAGES 36 & 96

NATURE.COM/NATURE

3 May 2015

Vol. 557, No. 7703

THIS WEEK

EDITORIALS

ETHICS Nature journals issue new policy on embryo and stem-cell work **p.6**

WORLD VIEW Psychology of pain needs more medical attention **p.7**



BIRD BEHAVIOUR How brothers of a feather can flock together **p.9**

Family connections

A high-profile arrest in California shows how the long arm of the law can now extend into DNA databases to check for relatives of suspected criminals.

Last week's arrest of a suspect in the Golden State Killer case in California has highlighted how DNA samples that have been volunteered for one purpose — in this case, genealogy — can be used for other reasons, often without the donor's explicit consent. Several ethicists have expressed concern about US detectives using a genealogy website in this way. Coming so soon after the reuse of Facebook data in political campaigns in the Cambridge Analytica scandal, it's another example of how new technology and techniques lead to unexpected conundrums, and how ethical and societal debate must catch up.

The case of the Golden State Killer, linked to at least 50 rapes and 12 murders between 1976 and 1986, had gone cold — although investigators believed they had a reliable sequence of the perpetrator's DNA. Next they needed a match. So, according to reports, they uploaded the data to a popular website that compares people's genetic information to trace their relatives — in effect, creating a profile for him. They got lucky: a match with family members led them to identify and arrest Joseph James DeAngelo.

Just like the Cambridge Analytica case, this one raises the question of how much control people have over information they give to public or commercial databases. DeAngelo's relatives submitted their DNA for the specific purpose of genealogy, which by definition requires the information to be shared and compared. Then they saw it used for something else without their consent. In discussions of the case, users of genealogy services are divided between those who say the police were justified, given the seriousness of the crimes, and those who were shocked by the move.

Such users have received other surprises. Thousands of people have discovered through genetic analyses that their parents were not who they thought they were. Others have found and been reunited with siblings they never knew existed. Such discoveries have implications for users' wider family members, most of whom won't have put their DNA in such a database.

In the California case, the involvement of the police adds an extra dimension. People who choose to upload their DNA could unknowingly be helping police to trace a relative — now and in the future.

Investigators have long coveted the genetic information held in others' databases. After the Swedish politician Anna Lindh was assassinated in 2003, Swedish police asked for access to a suspect's DNA stored in a biobank, so that they could compare it with DNA found at the crime scene. Their access was granted. But other requests have been turned down by courts. In 2006, the Norwegian Supreme Court said that police investigating a suspected armed robber, who had died six months after the crime, could not access his genetic information held by a hospital. The dead can't be libelled, but they can have their privacy invaded. And scientists in Belgium wrestled with these issues in 2016, when they confirmed the location of the 1934 death of King Albert I from blood samples collected there. They decided not to publish sequence details because of possible implications (including

paternity and health) for his surviving descendants, including members of the current Belgian and British royal families.

To what extent can scientists and companies who collect such information anticipate future uses and make them clear to participants and customers? There are no easy answers, because many of those uses cannot be anticipated at the time. Still, a 2016 survey showed that online

“Ethics should lead the conversation, rather than playing catch-up.”

firms that collate and compare DNA for consumers are too vague about how it might be used (E. Niemiec and H. C. Howard *Appl. Transl. Genom.* 8, 23–30; 2016). On the basis of what they do know, many organizations should take steps to inform people better. (In the California case, users of the site, GEDmatch, were told that “other uses” were

possible and that they should remove their data if this was unacceptable.)

If police can use genetic databases to catch killers — even those who are distant relatives of individuals who have submitted their DNA — then perhaps more people will sign up to share their DNA. But they should be told that this is a possibility, and be given the choice to opt out. Meanwhile, more geneticists, ethicists and lawyers need to debate other potential ways in which genetic information is likely to be used, so that ethics leads the conversation, rather than playing catch-up. ■

All that ails you

To improve health care, researchers need to study diseases as they occur: in combination.

When public-health researcher Tolullah Oni travelled from London to South Africa to study HIV, she soon realized she would have to broaden her focus. Physicians there were grappling with twin epidemics — HIV and tuberculosis. The infections often coincide, and so clinicians were working to integrate their treatment of the two diseases.

But Oni found that many of her patients were dealing with a third problem. “We started seeing people who came in with good adherence to their medicines, but somehow someone had missed the fact that their blood pressure was through the roof,” she says. To bring them back to health, she would need to treat non-communicable diseases such as high blood pressure and diabetes as well. “We were treating conditions and not people.” Oni went on to study the phenomenon in her patient community (T. Oni *et al. BMC Infect. Dis.* 15, 20; 2015) and is hoping to take the lessons learnt from integrating care of HIV and tuberculosis and apply them to other combinations of diseases.

People are complicated, and their medical problems rarely come neatly packaged as the single diseases that scientists and doctors study. A report released on 19 April by the UK Academy of Medical Sciences (see go.nature.com/2jhmvcvf) details the challenges of studying and treating individuals who have multiple medical conditions, known as multimorbidity. Variations in the definition and frequency of multimorbidity across populations have led to wide estimates of its prevalence, ranging from 13% to 95% of patients globally. The report offers a list of recommendations on what health-care providers can do to address the problem of multimorbidity, and identifies the knowledge gaps that need to be filled.

Researchers should take heed: if their work is to translate to the real world, more scientists — at the clinic and the bench — should shift their focus to look at interactions between disorders.

Multimorbidity seems to be growing in countries where the population is ageing and thus more people are living with chronic diseases, and in countries grappling with chronic infectious diseases such as HIV. Health-care providers should look again at how doctors tend to specialize in specific disorders, when it might be better to arm them with the ability to recognize and treat a range of conditions.

Clinical trials have historically focused on single diseases. They often exclude participants with other conditions to boost the chance of getting a cleaner data set (and to reduce risks of unintended harm). But this is beginning to change as part of a push to lower eligibility requirements for many clinical trials. Researchers are also increasingly focusing on supplementing data from carefully controlled clinical trials with ‘real-world evidence’ — much messier data collected from people who may be taking multiple medications and dealing with multiple conditions. Such studies are a good way to start understanding the effects of multimorbidity. In this issue, a World View describes how to make sure people with anxiety disorder and other complications

are integrated into clinical research of pain treatments (see page 7).

There is more to be done. As the report highlights, clinical researchers need to characterize multimorbidity around the world, looking at which conditions are most likely to coincide and in which populations. Already, evidence shows that this varies dramatically by location and wealth. More-deprived individuals in wealthy countries, for example, might be more likely to have multiple chronic diseases; whereas in poorer countries, wealthier individuals might be more likely to have multiple conditions.

“Multimorbidity seems to be growing in countries where the population is ageing.”

Such studies could identify the most prevalent and harmful clusters of disease — and so help to focus basic research. Bench scientists also tend to focus on one disease at a time, even if their work sometimes yields insights into a range of conditions. More effort should be put into studying complex combinations of disorders and how they — and their treatments — interact. Studies of ageing, for example, are detailing the causes of inflammation and its impact on multiple organs in the body (M. N. Bouchlaka *et al.* *J. Exp. Med.* **210**, 2223–2237; 2013).

This requires support from funders, and a wider recognition that the most tractable projects with the cleanest, easiest to interpret results might not be the most worthy of funding. Studying diseases in combination is challenging, but computational and laboratory tools are increasingly available to handle complex data sets and tease out meaning from messy data.

Some funders are already taking steps in this direction: an upcoming workshop held by UK charity the Wellcome Trust, the UK Medical Research Council and other organizations will look at how research can better tackle multimorbidity. This movement needs support in the coming years. Awareness of multimorbidity has been growing steadily: now the question is how best to deal with it. ■

ANNOUNCEMENT

Human embryo and stem-cell research

Research using human embryos and embryonic stem cells draws intense ethical scrutiny and places demands on scientists, funders and journals to follow the relevant regulations. As a publisher of such work, *Nature* and the *Nature* journals take this responsibility very seriously. For many years, *Nature* journal editors handling manuscripts on human embryo and stem-cell research have assessed the ethical oversight of the work when deciding whether to publish it. We are now formalizing and amending aspects of this publication policy.

Nature journals encourage stem-cell scientists to embrace guidelines agreed in 2016 by the International Society for Stem Cell Research (ISSCR) as they design, execute and report their research. These ‘Guidelines for stem cell research and clinical translation’ describe rigorous standards for stem-cell research consistent with international policies that govern biomedical science and clinical trials. To encourage scientists to follow these guidelines, we have identified categories of manuscripts for which we will require authors to send an accompanying ethics statement or will consult an ethicist reviewer.

Under this policy, *Nature* journals will require an ethics statement from the authors for papers that involve human embryos or gametes, and for clinical studies of cells derived from pluripotent stem cells. This statement must highlight ethical oversight of the work, including the review boards specialized in embryo research that approved

it, and details of the consent process for cell donors and recipients.

For manuscripts that we consider especially sensitive, *Nature* journals will request assessment by an independent ethicist alongside scientific peer review. Such manuscripts will include, but will not be limited to, those reporting genome engineering of human embryos or clinical work with gametes or cells derived from pluripotent stem cells. These ethicist reviewers may provide guidance on formulating the ethics statement to ensure accurate and transparent reporting of approval conditions. Authors may be asked to submit redacted informed-consent documents and review-board documents for evaluation by the ethicist reviewer.

Independent ethics review will also be required for manuscripts reporting work in which intact human embryos or embryo-like structures are kept alive for close to 14 days, a time point that corresponds to the formation of the primitive streak and the acquisition of organismal potential.

At present, many countries — and the ISSCR guidelines — prohibit culture beyond 14 days, a restriction that reflects the conclusions of the 1984 UK Report of the Committee of Inquiry into Human Fertilisation and Embryology (also known as the Warnock report). Whether this rule should be relaxed is currently being debated, triggered in part by technological advances that enable scientists to reconstruct human embryo-like structures from stem cells.

As this and other debates unfold, we anticipate the need to revisit some aspects of our policy in accordance with shifts in best practices for the stem-cell field, driven by advances in science and technology and evolving social norms. *Nature* fully supports an inclusive approach to such discussions, involving broad consultation and dialogue. We hope that our policy complements these efforts by scientists, ethicists, regulators, policymakers and funding agencies. ■

IAN MACKAY



To treat pain, study people in all their complexity

Clinical research needs to investigate not simply drugs, but the psychology of why and how individuals experience pain, says Beth Darnall.

Last month, the US National Institutes of Health (NIH) formally launched a multi-agency effort to combat the country's opioid-addiction crisis. Funds for research into controlling opioid misuse and treating pain will nearly double in 2018, to US\$1.1 billion.

The forces behind this epidemic extend beyond overprescription: most of the tens of thousands of deaths caused by opioid overdose in the United States each year result from illicit use. Still, an inadequate understanding about how to treat pain has certainly contributed. We need to characterize patients better, and we need more studies that incorporate non-drug treatments alongside any form of medication.

Consider this crucial question: what is the first treatment you should give a person for chronic pain, or even many acute injuries? Most clinicians now agree that the answer should not be opioids. Fewer recognize that the question is not which pill to use instead, but what system of interventions — including medication — and monitoring to implement.

Too often, pain is treated as a purely biomedical problem. It is a biopsychosocial condition. Psychological treatment can be combined with medication to equip people with the tools to better control their pain experience. Psychological therapies can also lower risks such as addiction, because the emphasis is on engaging patients in managing their daily actions to help themselves to feel better in the long run, rather than relying solely on passive medications. Yet a common clinical practice is to recommend such psychosocial strategies for pain only after all medications have failed.

It is hard for clinicians to learn which treatments to use, because our research system shuns the very patients we need to understand. Pain-research trials often exclude adults who have depression, anxiety and other disorders, those who take other prescription medications and those over the age of 70, who tend to have multiple co-morbidities.

To treat pain better, we should attend to these complex patients, rather than exclude them. One effort to do so is the Collaborative Health Outcomes Information Registry, or CHOIR (<http://choir.stanford.edu>), which my colleagues developed with NIH support. The platform collects data on the patients we personally see in our pain clinic every day: their age and sex; how they are sleeping; how pain affects their daily routine; their mobility, strength and endurance; how they engage with friends and family; which other medications they take; what other diagnoses they have. It also tracks treatments and responses over time. Clinicians can easily follow their own patients' progress, and the system can be programmed to recommend tailored treatments or patient education.

Researchers can look for patterns among groups of patients. For example, several studies suggest that most people taking opioids long-term do not benefit from them (see go.nature.com/2vyvlpk). However, almost all clinicians who treat chronic pain observe that some people do quite well on opioids. We need to be able to predict who those individuals are. Otherwise, we are either going to exclude people from a

treatment that benefits them or expose them to a risky medication.

We know, for instance, that people who worry more about pain, or who report feeling helpless in the face of it, are at risk of prolonged pain and opioid use after surgery (M. M. Wertli *et al. Spine* **39**, 263–273; 2014). My colleagues and I are currently assessing whether an online education app can help patients to manage their worries, decrease pain and limit opioid use after surgery.

More such pragmatic clinical trials are needed. So are accessible tools, such as CHOIR, to implement these trials. We are currently building a CHOIR network across the United States, Canada and Israel to integrate data and answer questions about which of several commonly used pain treatments works best, and in which individuals. Ideally, we will then use the results of these trials to inform clinicians continuously about the most safe and effective treatment to prescribe for their patients.

And we need to study how placebo effects could enhance pain treatment, by deliberately integrating them into clinical trials. I am not talking about sugar pills, but about a strategy called placebo optimization. Simple pain-science education, cognitive regulation and relaxation skills can help empower patients to reduce pain processing in the brain, gain better control over their symptoms and garner more benefit from medical treatments.

Patients can actually be primed for relief. For instance, placebo optimization could involve emphasizing to patients that we have evidence suggesting that various treatment plans — such as gently tapering opioid dosing — can be done without increased pain. Clinicians also need strategies for detecting and minimizing 'nocebo' effects: in this case, negative expectations and fears

about pain that can undermine the effectiveness of medical treatment.

We need to incorporate psychology and complexities into clinical trials and medical care. More funding for treating opioid addiction and misuse is welcome. But essential, too, are funds for investigating pain as a condition in itself.

In 2016, the Institute of Medicine estimated that up to one-third of the US population lives with ongoing pain. Chronic pain, the main cause of disability, is more prevalent than diabetes or heart disease. It costs the US economy up to \$630 billion every year in health care and lost productivity, and lowers the quality of too many lives. Although precise numbers are hard to come by, NIH spending breakdowns show that the agency committed just over \$500 million in 2017 to broad pain research. Finding better ways of treating pain is surely worth a greater investment. ■

Beth Darnall is a clinical professor in the Department of Anesthesiology, Perioperative and Pain Medicine at Stanford University in California.
Twitter: @bethdarnall

**TOO OFTEN,
PAIN IS
TREATED
AS A PURELY
BIOMEDICAL
PROBLEM.**

SEVEN DAYS

The news in brief

EVENTS

AI boost for Europe

The European Commission will increase its spending on artificial intelligence (AI) to €500 million (US\$604 million) per year for three years starting in 2018, up from about €300 million the previous year. The cash, announced on 25 April, is part of an initiative designed to boost Europe's standing in the field. Other plans include creating ethical guidelines for AI development and proposing legislation to increase the amount of publicly available data. Separately, a group of prominent European AI researchers signed an open letter on 24 April warning that the continent's AI laboratories, investments and companies are not keeping pace with rivals in North America and China. The statement calls on European governments to create an AI institute with sites in several countries, similar in scope to the European Molecular Biology Laboratory. Each location should have an initial investment of around €100 million, the letter says.

Facebook data

Facebook's tightening of third-party access to user data in the wake of the Cambridge Analytica controversy risks hampering research, a group of academics has argued. In an open letter published on 25 April, prominent data and Internet scientists said that restrictions on how third parties access social-media data were likely to diminish transparency and independent oversight of such platforms. They welcomed a Facebook initiative, announced on 9 April, to encourage peer-reviewed research on the role of social media in elections and democracy, but they said that the proposal's narrow terms of reference and use of a hand-picked panel of scholars

to define the research agenda mean that it risks failing to support independent research.

Suspected killer

Law-enforcement officials in California used DNA data found on a genealogy website to track down a suspected serial rapist and murderer, raising widespread concerns about genetic privacy. *The Sacramento Bee* newspaper broke the story on 26 April, reporting that investigators had used the free database, called GEDmatch, to find relatives of the suspected 'Golden State Killer', whose alleged crimes date back to the 1970s. In a statement, GEDmatch said that it had not been consulted about this use of its data, and noted that it had always warned users that the database could be used for

other purposes. The suspect, who is 72, has so far been charged with 8 murders. See page 5 for more.

POLICY

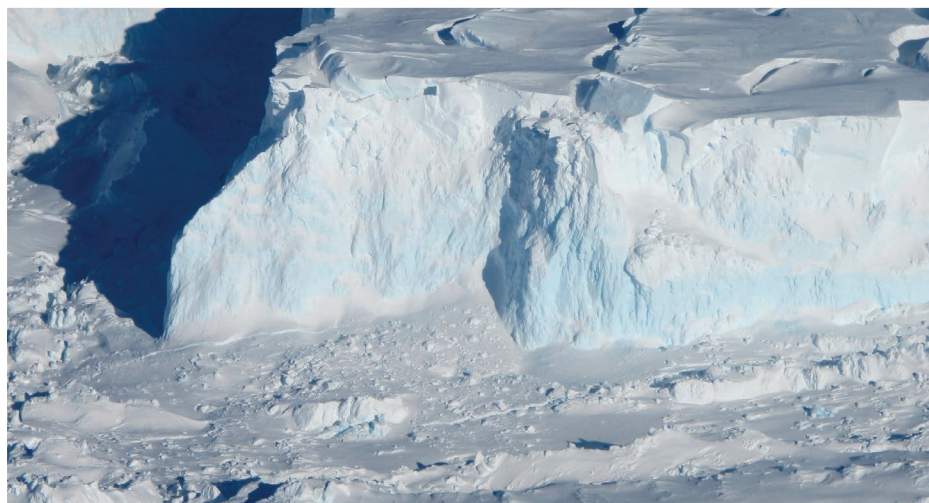
Letter to Trump

Nearly 700 members of the US National Academy of Sciences (NAS) have signed a public letter that denounces the administration of US President Donald Trump for its hostility to science. The statement, published on 23 April, admonishes Trump for withdrawing the United States from the Paris climate accord, and warns of the consequences of disregarding scientific evidence. The letter encourages the administration to maintain scientific content on publicly available websites, to appoint qualified people to posts

requiring scientific knowledge, to stop intimidating government researchers and to rejoin the Paris Agreement. The members signed as individuals, and not on behalf of the NAS.

Insecticide ban

The European Union has voted to ban the use of certain controversial neonicotinoid insecticides on all outdoor crops. The vote, which took place on 27 April, ends years of bitter wrangling between those supporting a ban, including environmentalists and many scientists, and opponents. An influential scientific review concluded in February that the insecticides pose a high risk to wild bees and honeybees. The three neonicotinoids of greatest concern for bee health — clothianidin,



JAMES YUNGEL/NASA

Nations join to watch for glacier collapse

UK and US polar scientists are launching a £20-million (US\$27-million) effort to probe the Thwaites glacier in Antarctica. The five-year project, announced on 30 April, is the biggest joint Antarctic effort by the nations in more than seven decades. The programme will fund eight studies and is set to begin this October. Researchers will gather radar, seismic and other

data on the glacier to understand whether it is headed for collapse. The glacier's drainage basin covers an area roughly the size of Britain on the West Antarctic Ice Sheet, and it already accounts for about 4% of global sea-level rise. The programme is being funded by the UK Natural Environment Research Council and the US National Science Foundation.

imidacloprid and thiamethoxam — will not be allowed to be used outdoors, but can be used in permanent greenhouses. The ban is binding for all EU member states, and it will enter into force by the end of 2018.

SPACE

Mars-rock return

NASA and the European Space Agency (ESA) are considering a joint mission to bring soil samples from Mars to Earth, a statement announced on 26 April. The venture, which would help to illuminate Mars's potential to harbour life, is no small feat. It would require both agencies' future Mars rovers, which are set to land on the red planet in 2021, to collect soil samples from the Martian surface and just beneath. A third rover would pick up the samples and place them in a rocket to be launched into a Martian orbit, where it would rendezvous with a spacecraft that would fetch the specimens and bring them to Earth. Plans for the rover and spacecraft are yet to be approved.

Moon mission

NASA has apparently cancelled a future lunar-rover mission, despite a directive from US President Donald Trump to focus on returning to the Moon. The Resource Prospector spacecraft was

scheduled to launch in 2022 to mine substances such as hydrogen and water. Results from that mission would have been used to inform human exploration of the Solar System. On 27 April, NASA head Jim Bridenstine tweeted that Resource Prospector's instruments would still be used, presumably separately from the rover, in missions to the surface of the Moon. One influential group of lunar scientists, the Lunar Exploration Analysis Group, criticized the move and argued that the rover should be a key component of NASA's renewed focus on lunar landings.

POLITICS

Nuclear site

North Korea's mountain nuclear test site (pictured) partially collapsed after the most recent nuclear detonation in September 2017, say Chinese researchers. A 4.1-magnitude earthquake occurred 8.5 minutes after the blast at Punggye-ri in the country's north, followed by several smaller earthquakes 20 days later; the authors say these tremors indicate rock falling in from above the blast cavity and note that the mountain should be monitored for potential radioactive leakage. The findings were published on 27 April (D. Tian *et al. Geophys. Res. Lett.* <http://doi.org/cn3t>; 2018), the same



day that North Korean leader Kim Jong-un met South Korea's President Moon Jae-in and pledged to close the nuclear test site in May.

Record wait

The US National Oceanic and Atmospheric Administration (NOAA) has been without a permanent administrator for more than 15 months — a record for the agency. US President Donald Trump nominated AccuWeather chief executive Barry Myers to head NOAA in October 2017, pending Myers's confirmation by the Senate. But the Senate vote has been held up over concerns about potential conflicts of interest. AccuWeather uses NOAA data to provide a host of weather-related services, and is owned and operated by Myers and his two brothers, Joel and Evan. Myers has said he will step down and divest himself of interests in the company

BUSINESS

CRISPR arguments

A US appeals court heard oral arguments in the ongoing dispute over rights to key CRISPR-Cas9 genome-editing patents on 30 April. The University of California and its collaborators are appealing against a 2017 decision by the US Patent and Trademark Office to recognize a competing patent filed by a group led by the Broad Institute in Cambridge, Massachusetts. In the appeal, the California team argued that the patent office erred in deciding that the Broad's CRISPR patent represented a significant invention beyond that covered by the University of California's patent. If either party is not satisfied by the appeals court's decision, expected later this year, they could then appeal to the US Supreme Court.

ENVIRONMENT

Reef rescue

Australia's government will spend around Aus\$500 million (US\$377 million) to help the beleaguered Great Barrier Reef, it said on 29 April. Aus\$444 million will go to the Great Barrier Reef Foundation to tackle threats such as water pollution and invasion by crown-of-thorns starfish, and to support restoration efforts. Another Aus\$56 million will be given to the Great Barrier Reef Marine Park Authority to expand its management of the reef. Critics pointed out that the funding ignores the reef's biggest threat — climate change. A study last month found that global warming was a factor in the 2016 coral-bleaching events that damaged around one-third of the reef's corals (T. P. Hughes *et al. Nature* **556**, 492–496; 2018).

TREND WATCH

Scientists have reason to hope that North Korea will soon open up to more collaborations after historic peace talks last week with South Korea. North Korea publishes little research, but its output is growing. Its scientists published about 80 articles in international journals last year, more than 4 times their 2014 output, according to the Web of Science database. Some 60% of North Korean papers since 2015 name Chinese co-authors. Main topic areas include geosciences, engineering and materials science.

NORTH KOREA'S SCIENCE

The isolated nation publishes just tens of articles in international journals each year. Its researchers' main collaborators are in neighbouring China.



NATURE.COM

For daily news updates see:

www.nature.com/news

NEWS IN FOCUS

PUBLIC HEALTH Pre-emptive antibiotic treatment helps children in Africa **p.14**

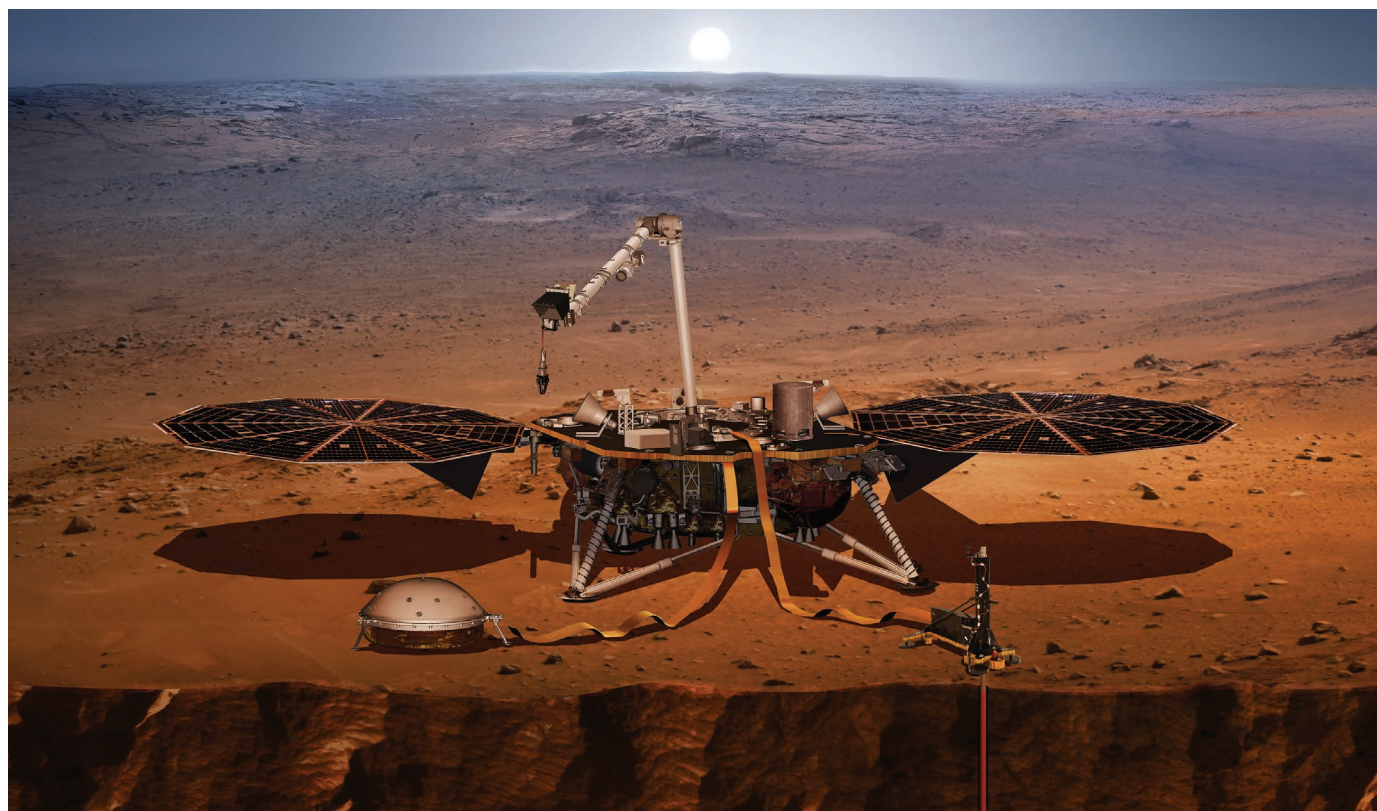
SYNTHETIC BIOLOGY Ambitious project to synthesize human genome shifts course **p.16**

ASTRONOMY Milky Way mapping mission releases huge data set **p.18**



WAR AND PEACE Understanding and conquering the roots of violence in Colombia **p.19**

NASA/JPL-CALTECH



An artist's impression of NASA's Mars InSight lander, which will touch down in November and listen for seismic tremors.

PLANETARY SCIENCE

Mars probe to hunt quakes

NASA's InSight mission will listen for seismic activity to uncover details of the red planet's mysterious core.

BY ALEXANDRA WITZE

A planetary stethoscope will soon be on its way to listen to Mars's heartbeat. On 5 May, NASA plans to launch its US\$994-million InSight spacecraft from Vandenberg Air Force Base in California. The mission's main job will be to place a seismometer on the Martian surface and listen to seismic waves pinging around the planet's interior.

If the effort succeeds, it will mark the first unequivocal detection of tremors known as marsquakes — and explain long-standing mysteries about the planet's inner structure

and how it evolved. "There are all these questions about Mars that can only be answered with seismic data," says Bruce Banerdt, a geophysicist at NASA's Jet Propulsion Laboratory in Pasadena, California, and the mission's principal investigator.

"It will be the first geophysical observatory on Mars," adds Ana-Catalina Plesa, a planetary geophysicist at the German Aerospace Center (DLR) in Berlin. "We are all really excited."

On Earth, seismologists use measuring stations scattered around the world to detect seismic waves from distant earthquakes. By tracking how that energy bounces around

the planet's interior, researchers can calculate information such as the size of Earth's core.

But no one has yet done this on Mars. NASA tried unsuccessfully with its Viking landers, which launched in 1975. Viking 1 failed to deploy its seismometer. And although the Viking 2 instrument gathered about 2,100 hours of data, all the tremors it detected, with one possible exception¹, were caused by gusts of wind shaking the spacecraft. The seismometer had been mounted on top of the lander rather than in direct contact with Mars's surface.

After InSight lands, it will plop its watermelon-sized seismometer — built by the ▶

► French space agency CNES — onto the Martian ground. The instrument will nestle beneath a protective wind shield as its three delicate pendulums measure even tiny tremors. “It is pretty much the most sensitive seismometer that’s ever been built,” says Renee Weber, a planetary scientist at NASA’s Marshall Space Flight Center in Huntsville, Alabama.

MARTIAN MYSTERY

The big question is how many marsquakes it will capture. With no actual data on Martian seismicity, researchers have used maps of geological faults on the planet’s surface², along with calculations of how its interior has cooled over time³, to estimate that Mars probably has fewer quakes than Earth but more than the Moon. (Tectonic fractures and Earth’s tidal pull trigger moonquakes.)

InSight will land in Elysium Planitia, a safe, flat and geologically boring site near the Martian equator⁴. There, it might measure one local marsquake each year between magnitude 2.7 and 4.2, says Weber. It could

also detect bigger marsquakes from regions much farther away, such as the fault-riddled Cerberus Fossae. “Our goal is to collect something like 30 quakes over the mission,” says Philippe Lognonné, a geophysicist at the Paris Institute of Earth Physics who leads the seismometer team.

The bigger the marsquake, the more it will reveal about the planet’s interior, because only the largest seismic events penetrate all the way to the core. InSight might see one or two such quakes during the two Earth years that NASA hopes to operate the mission.

Marsquake data will help InSight to map the boundaries between Mars’s crust, mantle and core. That could reveal the depth to which the planet’s primordial magma ocean once churned, and whether Mars ever had anything resembling plate tectonics. Pinning down the size of the Martian core, thought to be roughly half as big as Earth’s, would reveal its density and composition⁵. Mars’s internal layers represent a record of its first tens of millions of years, says Banerdt. And studying its interior might

help to reveal the early history of Earth, which probably went through similar changes soon after it formed.

Meanwhile, a radio-science experiment on InSight will measure how Mars wobbles on its axis, as a way to further understand the core’s size. And a heat-flow probe, built by a team at the DLR, will penetrate up to 5 metres beneath the surface to measure how temperature changes with depth and time.

The plan is to land InSight on Mars on 26 November. For Lognonné, who has been trying to get a seismometer to Mars for more than two decades, that day can’t come soon enough.

“I’ll be much more happy when I get the first data,” he says. ■

1. Lorenz, R. D., Nakamura, Y. & Murphy, J. R. *Earth Space Sci.* **4**, 681–688 (2017).
2. Knapmeyer, M. et al. *J. Geophys. Res.* **111**, E11006 (2006).
3. Plesa, A.-C. et al. *Geophys. Res. Lett.* **45**, 2580–2589 (2018).
4. Golombek, M. et al. *Space Sci. Rev.* **211**, 5–95 (2017).
5. Hellfrich, G. *Prog. Earth Planet. Sci.* **4**, 24 (2017).

PUBLIC HEALTH

Trial helps African children

Pre-emptive antibiotic treatment reduces deaths in at-risk kids, but raises fears about the development of drug resistance.

BY AMY MAXMEN

To stem the rise in antibiotic resistance, researchers recommend that people take these drugs only after they are diagnosed with a bacterial infection. But a trial involving nearly 200,000 children in Niger, Tanzania and Malawi went against that guidance in an attempt to save youngsters in regions where as many as one in ten die before their fifth birthday.

The results, published on 25 April in *The New England Journal of Medicine*, suggest that widespread distribution of antibiotics could prevent thousands of deaths (J. D. Keenan et al. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa1715474>; 2018). But health officials and researchers are wary of starting a massive drug

programme on the basis of the results because it would drive antibiotic resistance.

“It goes against dogma at the moment because everyone else is trying to reduce antibiotic use,” says Per Ashorn, who specializes in paediatric infectious diseases at the World Health Organization (WHO). The study’s results are exciting, he says, but the WHO needs more data to evaluate the approach.

Some officials sound more enthusiastic about the strategy. “As a person who was born in one of the poorest countries in the world, I welcome this,” says Samba Sow, the health minister in Mali, where 11% of children die before the age of 5. “My older brother died as a child, more than one of my cousins died as a child — children die here, and they die fast.”

Mass drug administration fell out of favour in the late 1960s, after programmes to prevent malaria through the large-scale distribution of a drug called chloroquine backfired. Resistance developed rapidly because thousands of people with insufficient levels of chloroquine in their systems became infected with malaria parasites. Strains that were susceptible to the drug died, but more-resilient ones multiplied. Eventually, chloroquine stopped working.

But opinions on mass drug administration seem to be changing. Since 2012, several African countries have reduced deaths from malaria by pre-emptively treating millions of children during the rainy season. And the WHO now recommends the distribution of some antibiotics to certain populations impacted by neglected



TOP NEWS



How easing relations with North Korea might bolster research
go.nature.com/2hepjg

MORE NEWS

- Harassment and discrimination allegations roil top US biomedical institute go.nature.com/2jhzrwg
- Europe bans bee-harming pesticides go.nature.com/2jgqjq5
- Huge South Korean quake linked to geothermal plant go.nature.com/2jhmrvv

NATURE PODCAST



Constructing early embryos; how mice react to danger; and an ancient butchered rhino nature.com/nature/podcast

RAYMOND CUNNINGHAM/GETTY

tropical diseases — a constellation of illnesses affecting roughly one billion people living in poverty around the world.

The idea for the latest study came from an analysis of the pre-emptive use of the antibiotic azithromycin in Ethiopian communities affected by trachoma — a disease that causes blindness — in the late 2000s. Researchers noticed a drop in overall deaths (T. C. Porco *et al. J. Am. Med. Assoc.* **302**, 962–968; 2009), and Thomas Lietman, an infectious-disease researcher at the University of California, San Francisco, and his colleagues followed up with the current trial, dubbed MORDOR (from the French description of the project).

As part of MORDOR, children under five in communities in Niger, Malawi and Tanzania took one dose of azithromycin twice a year for two years. Control populations received a placebo. Childhood mortality rates among treated communities in Niger dropped by 18% compared with control populations; Tanzania had 3% fewer deaths and Malawi saw a 6% reduction.

Lietman says that Niger probably experienced the greatest benefit because it has the highest childhood mortality rate of the three countries. About 9% of children die before the age of 5 in Niger, compared with about 5% in Tanzania and Malawi. Pneumonia and diarrhoea triggered by bacterial infections help to drive up childhood mortality rates, according to a 2017 report from the United Nations. Poor sanitation, unsafe drinking water and malnutrition combine to make children living in poverty especially vulnerable to disease-causing microbes. They're also more likely to die from curable conditions because health care can be unaffordable or too far away to be of help.

But this antibiotics strategy comes at a cost, says Ramanan Laxminarayan, director of the Center for Disease Dynamics, Economics and Policy in Washington DC. If resistance develops against azithromycin, diseases treated by the drug, including gonorrhoea, would become harder to combat. He hopes that, if policymakers decide to implement this approach, they will target only the populations most in need, and then just for a limited time. Groups supporting this approach should also work to reduce childhood mortality in the same way that the developed world did, he says, through improved sanitation, nutrition and health care.

For now, researchers will continue to study the effects of this antibiotics strategy. Later this year, similar trials will launch in Mali and Burkina Faso. And Lietman's team is evaluating data collected during an extension of its trial, to assess how fast antibiotic resistance develops. The WHO plans to release a statement about whether this strategy is justified, and in what circumstances, by the end of 2019. ■



The Owens Valley Long Wavelength Array in California hosts the LEDA experiment.

ASTRONOMY

Physicists trawl skies for enigmatic signal

Teams rush to find faint signature from Universe's first stars.

BY DAVIDE CASTELVECCHI

Researchers are heading to some of the most remote spots on Earth — from the Tibetan Plateau to an island in the sub-Antarctic ocean — to try to capture an enigmatic radio signal from the early Universe. This grand search includes some of the first experiments to follow up on a surprise announcement in February that astronomers had seen evidence of the Universe's first stars lighting up¹, a moment known as the cosmic dawn.

And as experimental physicists try to replicate those findings in the few places on Earth that are relatively undisturbed by radio interference, theorists are struggling to make sense of the data. "The signal does not look like anything we expected," says Abraham Loeb, an astrophysicist at Harvard University in Cambridge, Massachusetts.

The original detection was reported by researchers at the Experiment to Detect the Global Epoch of Reionization Signature (EDGES), which uses a pair of table-sized radio antennas in the Australian outback. The experiment measures the long-wavelength part of the cosmic microwave background, the noisy afterglow of the Big Bang. The researchers were searching for a subtle dip in the background spectrum where the microwave radiation is

slightly dimmed. Cosmologists have theorized that such a dip should have been caused by the light of the first stars, which made primordial hydrogen in the Universe less transparent at a particular radio wavelength. The details of this absorption should contain information about the early interstellar matter and the stars that cast light on it.

"We might have the most radio quiet place on Earth."

But the blip had an unexpected shape. It suggested that the absorption started to ramp up rapidly

around 150 million years after the Big Bang, stayed roughly constant between 200 million and 250 million years ago, and then disappeared relatively quickly. The dip was also deeper than predicted, which implied that the gas was colder than expected during that epoch — perhaps 4 kelvin instead of 7 kelvin.

EXTRA SCRUTINY

The EDGES team spent two years cross-checking their peculiar result before deciding to go public. Researchers have posted dozens of preprints in response, trying to interpret the anomaly. Some physicists have suggested that it was a possible sign of previously undiscovered interactions between ordinary matter and dark matter². Others saw a possible indication of ►

► the absence of dark matter³.

The EDGES researchers have now started another round of observations with a new, smaller antenna. They have “preliminary evidence” that this antenna also sees the original feature, says lead scientist Judd Bowman, an astronomer at Arizona State University in Tempe.

Competing experiments are also trying to reproduce the EDGES result. In April, Lincoln Greenhill, a radio astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, flew to the arid Owens Valley in California to test a modified version of the Large-Aperture Experiment to Detect the Dark Ages (LEDA). The experiment — an array of antennas that look like umbrella frames — might have just missed the EDGES signal because it originally operated with filters that cut off frequencies above 82 megahertz. The EDGES signal seems to be centred at about 78 megahertz, so is very near the top of that range. The team is testing filters that allow detection of higher frequencies. If things go well, Greenhill says, it might take a few months to collect and analyse enough data.

Meanwhile, at the Raman Research Institute in Bangalore, India, Ravi Subrahmanyan and his colleagues have quickly built a version of their spherical antenna, called SARAS-2, that is sensitive to the range of the EDGES signal. They plan to deploy the new antenna in May at a site outside town, and to later move it to the Tibetan Plateau.

Places without radio interference are rare now, but “we might have the most radio quiet place on Earth”, says physicist Jonathan Sievers at the University of KwaZulu-Natal in Durban, South Africa. The spot is on Marion Island, halfway to Antarctica, and the only way to get there is on a ship that goes once a year, in April. A small KwaZulu-Natal team led by physicist Cynthia Chiang installed its cosmic-dawn experiment, Probing Radio Intensity at High-Z from Marion (PRIZM), there last year. Chiang is now at the island station again, retrieving data from the past year and upgrading their telescope for new observations.

But one of the quietest places for radioastronomy in the Solar System would be the far side of the Moon. Jack Burns, an astrophysicist at the University of Colorado Boulder, is leading a proposal to put a 10-metre-long wire antenna on a small lunar orbiter. From there, the probe should detect not only the EDGES absorption feature, but also one from an earlier epoch known as the dark ages — before stars existed. The feature would appear at around 15 megahertz, a band that is not accessible from Earth. ■

1. Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J. & Mahesh, N. *Nature* **555**, 67–70 (2018).

2. Barkana, R. *Nature* **555**, 71–74 (2018).

3. McGaugh, S. Preprint at <https://arxiv.org/abs/1803.02365> (2018).

SYNTHETIC BIOLOGY

Genome-synthesis effort shifts focus

GP-write project to make virus-resistant human cell lines.

BY ELIE DOLGIN

A bold plan to synthesize an entire human genome has been scaled back to a more technically attainable near-term goal. Instead of synthesizing all of the human genome’s 3 billion DNA base pairs, the project will now attempt to recode the genome to produce cells immune to viral infection.

Organizers of Genome Project-Write (GP-write), a global public-private partnership that includes around 200 scientists, announced the priority shift at a meeting in Boston, Massachusetts, on 1 May.

But even the downsized ambitions might be difficult to achieve soon, because the two-year-old effort still has no dedicated funding for what’s estimated to cost tens, if not hundreds, of millions of dollars and last a decade or more.

“We thought it was important to have a community-wide project that people could get behind,” says project co-leader Jef Boeke, a yeast geneticist at New York University. When the effort launched in 2016, the creation of a virus-resistant human cell line was listed as one of several pilot projects that would develop the technology to synthesize the full genome¹. With the cell line now the focus, raising money should be easier, says Nancy Kelley, a biotechnology lawyer who is co-leading the effort with Boeke and George Church, a genome scientist at Harvard Medical School in Boston.

Onlookers generally approve of the priority shift. “This is a terrific idea,” says Martin Fussenegger, a synthetic biologist at the Swiss Federal Institute of Technology in Zurich. “It’s more geared toward utilities and applications” — not just DNA synthesis for its own sake, he adds.

A virus-proof human cell line would let firms make vaccines, antibodies and other biological drugs without risk of viral contamination. It could also help to make protein drugs with chemical ornaments similar to those in human proteins, to decrease the risk of the body’s immune system rejecting them. However, the organizers’ main goal is still to improve DNA technologies, not to create a particular product.

“The idea is to develop the technologies to

do this very quickly and easily using a variety of gene-editing and synthesis techniques,” says Harris Wang, a synthetic biologist at Columbia University Medical Center in New York City, and a member of GP-write’s scientific executive committee. The “ultra-safe” human-cell-line project, Wang adds, has “the right level of complexity, difficulty and many different facets of design” to push those technologies forward.

One thing it doesn’t have going for it, however, is much dedicated funding. Although a gene-editing technology company said it would donate technical expertise at the meeting, no financial backers have stepped forward.

Church estimates that the consortium has more than US\$500 million in “related funding” — but he includes, for instance, \$40 million earmarked for his own work on synthetic-biology projects including engineered bacteria and miniature organ-like structures. He also counts \$23.4 million for an international initiative led by Boeke to synthesize the yeast genome. Both efforts started years before GP-write.

And the lion’s share of the related funding is investment money raised by loosely affiliated biotech companies. Church includes it in his estimates not because the firms have given money to the effort, but because he is tabulating what he calls “a rough-draft market summary” of the gene-synthesis “ecosystem”.

As such, he includes hundreds of millions of dollars collectively raised by eGenesis, a start-up that he co-founded in Cambridge, Massachusetts; Twist Bioscience in San Francisco, California, of which he is a shareholder; and Ginkgo Bioworks, a Boston synthetic-biology company that last year acquired another Church-backed venture, Gen9. And although leaders of eGenesis and Twist have been active in GP-write, Ginkgo senior management has not. “We’re not involved in GP-write at all, and I’m surprised to see that they included us on that list of funding,” says creative director Christina Agapakis.

Church defends his accounting. “It would be great if we accomplish the goals of GP-write entirely with pre-existing or unlabelled funds,” he says. “Companies like Ginkgo are relevant independent of their formal ties.”

When (and if) the consortium can secure funding for its ultra-safe human-cell-line project, the team plans to imitate previous efforts by Church’s lab to recode the genome of *Escherichia coli* bacteria, making it resistant to viruses.

“It would be great if we accomplish the goals of GP-write entirely with pre-existing or unlabelled funds.”

In that project², researchers swapped all 321 instances of one 3-letter genetic word, or codon, with another that conveys the same message. They then eliminated the gene that allowed the cell to read the original codon. This didn't much affect the redesigned microbe, but it did neutralize viral invaders because, like all natural life, they rely on that codon for proper protein assembly.

Extending this recoding technique to the human genome won't be easy. Repurposing

just one codon across all 20,000 human genes will require hundreds of thousands of DNA changes. It might be easier to synthesize large swathes of the genome rather than edit letters one by one.

Church's team used synthesis in follow-up work³ to recode seven codons in the *E. coli* genome. That effort needed close to 150,000 genetic changes, and it revealed unexpected design constraints and difficulties in stitching together DNA fragments. These have

stymied efforts to make the reconstructed bacterium viable.

That should be a sobering reminder as the ultra-safe human-cell-line project gets off the ground, says Nili Ostrov, a postdoc in Church's lab who is leading the research. "In humans," she says, "there are going to be a lot of design rules that we just don't know." ■

1. Boeke, J. D. *et al. Science* **353**, 126–127 (2016).
2. Lajoie, M. J. *et al. Science* **342**, 357–360 (2013).
3. Ostrov, N. *et al. Science* **353**, 819–822 (2016).

ENVIRONMENT

Brazil's lawmakers push to weaken environmental rules

Legislation includes proposals to open up the Amazon rainforest to agriculture.



Trees taken in illegal logging operations in the Brazilian Amazon lie in piles at a sawmill.

BY JEFF TOLLEFSON

A conservative coalition that dominates Brazil's Congress is girding itself for a final push to roll back environmental regulations before campaigns for the country's October presidential election ramp up.

The legislation under consideration includes proposals to open up the Amazon rainforest to sugarcane farming — which was banned in 2009 over concerns about deforestation. Another proposal would weaken licensing requirements for infrastructure such as dams, roads and agricultural projects. But the rural-agricultural coalition behind the proposals is running up against public opposition that has thwarted previous efforts to loosen environmental rules.

Further complicating this fight is an ongoing corruption scandal that has landed former president Luiz Inácio Lula da Silva in jail. He was a leading candidate in this year's election before his conviction.

"There's this very delicate balance," says Mercedes Bustamante, an ecologist at the University of Brasília. The conservatives have support from Brazilian president Michel Temer as well as the votes they need to move legislation through Congress, she says. Lawmakers could push forward, Bustamante adds, but they're wary about sparking a public backlash before the election.

Previous efforts to scale back protected areas and indigenous rights in the Amazon rainforest floundered as activist groups and celebrities mobilized public opposition.

The conservatives have had only one major success on the environmental-regulation front so far. In 2012, they revised the Brazilian law governing forests, making changes such as eliminating penalties for any illegal deforestation that took place in the Amazon before July 2008. Environmental groups challenged the constitutionality of the revised law, but in February Brazil's Supreme Court upheld those changes.

"It was the worst thing that could have happened," says Carlos Nobre, a climate scientist in São José dos Campos and former secretary for research and development at Brazil's Ministry of Science, Technology and Innovation. But he thinks the conservative coalition's broader environmental agenda has stalled and is unlikely to advance in the coming months.

Brazil was once seen as a global leader on environmental issues, in large part because of its success in curbing deforestation. Between 2004 and 2012, the annual amount of rainforest that was cleared for agriculture fell by nearly 84% to 4,571 square kilometres. Those numbers subsequently crept back up, peaking at 7,893 square kilometres cleared in 2016. However, deforestation dropped by 16% to 6,624 square kilometres in 2017, partly because of lower demand for beef and the restoration of law-enforcement funding, which had been cut during a prolonged financial crisis.

The environment will certainly be on the election agenda, says Bustamante, because Lula's first environment minister, Marina Silva, is one of the candidates.

Regardless of the outcome, the political dynamic in Brazil's Congress is unlikely to change, says Paulo Barreto, a senior researcher with the activist group the Amazon Institute of People and the Environment in Belém. The conservative coalition is strong, and Barreto thinks that it will stay in power. ■

SPACE

Billion-star map set to transform astronomy

European Gaia mission releases most detailed 3D chart yet of Milky Way.

BY DAVIDE CASTELVECCHI

After a feverish wait, astronomers around the world have an ocean of new information to explore. On 25 April, the European Space Agency's (ESA) Gaia mission published its first fully 3D map of the Milky Way.

The data haul includes the positions of nearly 1.7 billion stars, and the distances, colours, velocities and directions of motion of about 1.3 billion of them. Together, they form an unprecedented live video of the sky, covering a volume of space 1,000 times larger than that captured by any previous survey (see 'Gaia's gold'). "In my professional opinion, this is crazy awesome," says Megan Bedell of the Center for Computational Astrophysics in New York City, one of the many astronomers who are already conducting studies based on the data set. "We're very curious to see what the community will do with it," says Anthony Brown, an astronomer at the Leiden Observatory in the Netherlands who chairs Gaia's data-processing collaboration.

At an event at the Royal Astronomical Society in London to present the Gaia catalogue, astronomer Gerry Gilmore of the University of Cambridge, UK, showed a striking video that

extrapolated from the Gaia data to simulate the future motions of millions of stars. "Everything moves," he said.

The 2-tonne Gaia spacecraft, part of a €1-billion (US\$1.2-billion) mission, launched in late 2013 and began collecting scientific data in July 2014. Gaia is in a stable orbit that remains fixed relative to both the Sun and Earth, and makes repeated measurements to estimate the distances to stars — and other celestial objects — using a technique called parallax.

Alongside its 551-gigabyte database, the Gaia team also released a number of scientific papers. The goal of these was to describe quality checks the researchers did on the data and demonstrate how those data can be used; the mission's policy is to make the catalogue immediately available to the community, rather than to reserve it for the team's own science studies first.

Still, the Gaia papers describe a wealth of original findings, said Floor van Leeuwen, another senior Gaia scientist at Cambridge, at the press briefing. He showed, for example, how Gaia proved for the first time that certain star clusters puff up at the same time as large stars sink to their centres. "We weren't allowed to make discoveries, but we couldn't avoid making them," he said.

One of those findings has implications far beyond the Milky Way. Some astronomers are eager to see Gaia's measurements of certain types of variable star that are used as 'standard candles' of cosmology. Knowing the precise distances to these stars in the Milky Way makes them useful as yardsticks for measuring distances to galaxies much farther away. In particular, astronomers use standard candles to estimate how fast the Universe is expanding, but in recent years, measurements based on this technique have been in apparent contradiction with predictions made using maps of the cosmic microwave background, the afterglow of the Big Bang. A preliminary look at the data shows that Gaia has improved the precision of the standard-candle measurements, Gilmore said at the press briefing. But, he adds, "at face value, the tension is still there".

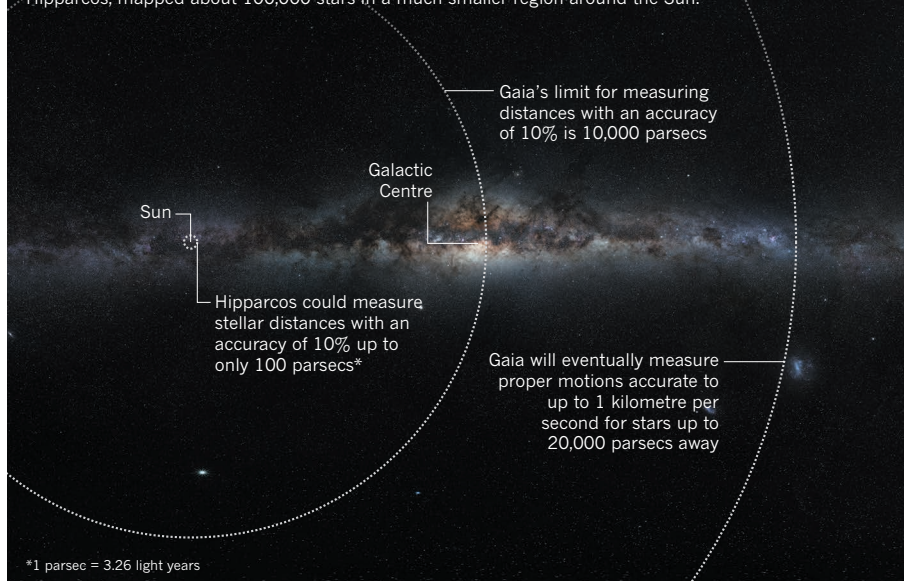
Dozens of preprints appeared in the days that followed, as teams around the world downloaded Gaia data and ran them through algorithms honed for years in preparation. For example, researchers are now able to test models of how the Milky Way formed through mergers of smaller galaxies and measure the distribution of dark matter.

Gaia released a preliminary catalogue in 2016, but at that time, it had not yet gathered enough data to directly measure the distances to many stars. Further data releases will contain more information and will enable entirely new kinds of studies; the next release will be in 2020. The probe also monitors asteroids and will help scientists to track bodies that look to be on a collision trajectory with Earth.

Gaia has enough fuel to keep operating until 2024, if nothing breaks down and ESA extends the mission beyond its current 2019 end date, says project scientist Timo Prusti at ESA's space-research and technology centre in Noordwijk, the Netherlands. The probe is in overall good health, he says. ■

GAIA'S GOLD

Gaia has measured with high precision the positions, distances and motions of more than 1 billion stars in the Milky Way. It covers about one-quarter of the disc of our Galaxy; its predecessor mission, Hipparcos, mapped about 100,000 stars in a much smaller region around the Sun.



CORRECTION

The News Feature 'The cells that sparked a revolution' (*Nature* **555**, 428–430; 2018) incorrectly stated that ViaCyte had restarted a 2014 clinical trial after redesigning its encapsulation technology. It had in fact paused enrolment on one trial and started another.



Tras la violencia

Los esfuerzos de paz en Colombia han puesto fin a 50 años de intenso conflicto. Ahora, científicos estudian a quienes fueron combatientes y a víctimas, al tiempo que ambos intentan sanar.

POR SARA REARDON

TRADUCIDO POR DEBBIE PONCHNER

Cuando empezó a estudiar a las personas que habían aterrorizado a su país, Natalia Trujillo se preparó para encontrarse cara a cara con monstruos.

Entrevistaría a excombatientes del largo y sangriento conflicto que se había apoderado de Colombia por más de 50 años. La compleja lucha de poder entre los insurgentes guerrilleros, el gobierno, los grupos paramilitares y los narcotraficantes había matado a cientos de miles de

personas y había desplazado a millones. Cuatro miembros de su familia habían sido secuestrados y la violencia había expulsado a su padre de sus tierras. Algunos de sus colegas habían pasado por experiencias mucho peores.

Trujillo, ahora una experta en neurociencia en la Universidad de Antioquia, en Medellín, estaba interesada en estudiar las raíces psicológicas de la violencia, observando a los combatientes que habían depuesto las armas e intentaban reintegrarse a la sociedad civil. Su oportunidad

llegó en 2010, cuando un programa de reintegración del gobierno reunió, por el día, en el jardín botánico de Medellín, a cientos de excombatientes.

Ella y su equipo de investigación ingresaron al enclave con una batería de pruebas cognitivas, botones de pánico –en caso de que algo saliera mal– y algunas ideas preconcebidas. “Pensé que personas que pueden matar a sus vecinos, que

Los combatientes de las FARC tienen la oportunidad de reintegrarse a la sociedad colombiana.



El municipio de Vista Hermosa podría servir como un laboratorio natural para el estudio de la reconciliación.

► pueden destruir sus comunidades, que pueden tener el corazón para obligar a otras personas a abandonar sus fincas, tienen que ser realmente malas”, dice Trujillo. Y se encontró con algunos que llenaron sus expectativas.

Con cadenas alrededor de sus cuellos y jactanciosos pavoneos, algunos trataron de intimidar a los investigadores. Pero con mayor frecuencia, los científicos encontraron personas comunes, paseando en el jardín y comiendo helado con sus hijos.

“Al principio estaba bastante decepcionada”, dice. Si algo andaba mal en sus cerebros, eso proporcionaría una explicación fácil para toda la maldad que habían hecho. Pero después de estudiar a más de 600 combatientes, ella comenzó a comprender la complejidad de sus experiencias. “Me di cuenta de que no todos son sociópatas. Me di cuenta de que la mayoría de ellos también son víctimas”.

Ese reconocimiento ha llevado a Trujillo y a sus colegas a reexaminar no solo sus propios sentimientos acerca de los excombatientes, sino también el enfoque que debe tener el país para lidiar con ellos.

El gobierno de Colombia participa actualmente en uno de los mayores esfuerzos de paz en la historia. Como parte de un tratado de 2016 con el grupo guerrillero de izquierda conocido como las Fuerzas Armadas Revolucionarias de Colombia (FARC), el gobierno otorgará amnistía a los combatientes que abandonen el conflicto y completen un programa de reincorporación, siempre que no hayan cometido delitos graves. Unos 6.800 combatientes de las FARC ya han ingresado al programa.

El esfuerzo, que es políticamente controvertido y se espera que cueste 129,5 billones de pesos colombianos (46 millones de dólares estadounidenses), enfrenta abrumadoras dificultades. Pero también le otorga a los

“Me di cuenta que no todos ellos son sociópatas. Me di cuenta que la mayoría de ellos también son víctimas”.

científicos una oportunidad única para comprender a una población que tanto ha infligido y como ha sufrido los horrores de la guerra. La mayoría de las investigaciones sobre las raíces psicológicas de la violencia y el trauma se han realizado con veteranos de países adinerados que lucharon en conflictos lejos de casa. La mayoría de los excombatientes de Colombia, por el contrario, tienen poca educación y están tratando de reingresar a la misma sociedad que una vez aterrizaron.

Allí enfrentan un enorme estigma y resentimiento, lo que les dificulta encontrar trabajo y entablar relaciones con los demás.

Un puñado de científicos están estudiando a los excombatientes con un detalle sin precedentes, con la esperanza de poder informar y guiar el proceso de paz. Han descubierto que los años de aislamiento y exposición a la violencia podrían haber alterado la psicología y el procesamiento cognitivo de los excombatientes de maneras sutiles. En pruebas de laboratorio, muchos tienen dificultades para identificarse con los demás y emiten juicios éticos errados –deficiencias que podrían afectar la forma en que participan en la vida civil–.

Los científicos ahora están iniciando

estudios a largo plazo en pueblos que estuvieron plagados por conflictos, para rastrear cómo la cognición y las actitudes podrían cambiar a lo largo del proceso de reconciliación, tanto para los excombatientes como para los civiles. Los datos podrían eventualmente informar los esfuerzos de recuperación de otros países devastados por la guerra.

Pero la investigación también está revelando cuán profundo es el desafío. Y algunos expertos temen que la atención disponible para los excombatientes, mientras tanto, es inadecuada.

“Va a ser increíblemente difícil salir de este círculo vicioso”, dice Jiovani Arias, psicoterapeuta y politólogo de la Universidad de los Andes, en Bogotá. Sin inversiones para mejorar la salud mental, dice, el legado de violencia que afecta tanto a excombatientes como a civiles podría torpedear los precarios esfuerzos de paz de Colombia.

EL CAMINO HACIA LA PAZ

En el momento que un autobús lleno de científicos llega al municipio de Vista Hermosa, en el centro de Colombia, Diana Matallana, neuropsicóloga de la Pontificia Universidad Javeriana, en Bogotá, todavía no puede creer en dónde está. “Hace cinco años, no podías venir aquí”, dice. “Es como un símbolo de la parte más dura del conflicto”.

El conflicto armado colombiano creció y menguó durante medio siglo mientras varios grupos militares competían por el control del territorio. Los civiles quedaron atrapados en el fuego cruzado: más de 260.000 personas murieron y 7 millones fueron desplazadas durante las décadas de violencia, según un registro gubernamental de víctimas.

La región del Meta, donde se encuentra Vista Hermosa, fue una de las muchas áreas abandonadas por el ejército colombiano en la década de 1990 y dejada para ser gobernada alternativamente por grupos paramilitares y guerrillas. Fue un arreglo tenso. Las guerrillas ayudaron a desarrollar infraestructura, pero no dudaban en matar rápidamente a informantes sospechosos. Los paramilitares, contratados principalmente por los capos de la droga y las élites políticas adineradas, tendían a ser más despiadados, torturaban a supuestos espías y dejaban cadáveres en las puertas de las escuelas. Ambas partes estaban fuertemente involucradas en el tráfico de cocaína y secuestraron a miles de personas a cambio de una recompensa –entre ellos, al hermano de Matallana–.

Desde el acuerdo de paz de 2016, se le ha permitido a los combatientes de las FARC entrar en una campaña de desarme y rehabilitación dirigida por la Agencia de Reincorporación y Normalización de Colombia (ARN), en Bogotá. La ARN había sido establecida años antes, y desde entonces ha facilitado la reintegración de unos 20.000 paramilitares y guerrilleros que abandonaron el conflicto de forma independiente o como parte de otro acuerdo de paz.

La ARN ahora opera 26 asentamientos

improvisados en toda Colombia, conocidos como zonas de tránsito (ver 'Un conflicto persistente'), para los miembros de las FARC que recién están reingresando a la sociedad. Ofrecen servicios como educación y atención de la salud, también ayudan a proporcionar algo de protección para los excombatientes, que son blanco habitual de antiguos enemigos, de antiguos aliados que se niegan a rendirse, y de civiles. Después de completar un programa, los excombatientes pueden recibir cédulas de identidad que les permite vivir y trabajar legalmente en el país.

La posibilidad de su regreso no emociona a algunos de los residentes de Vista Hermosa. Un letrero en la carretera que dice "Unidos, la paz y el pos-conflicto son posibles" ha sido bombardeado con bolas de pintura rosa. "Alguien no está de acuerdo", comenta uno de los investigadores.

Matallana y Carlos Gómez, un psiquiatra de la Javeriana, planean iniciar un estudio de 10 a 20 años de duración en el que se seguirá a más de 2.000 personas de Vista Hermosa, civiles y excombatientes por igual. "Estamos planeando por primera vez – en Colombia y en el mundo – aprender qué cosas ayudan a lograr la reconciliación", dice Gómez.

El equipo pretende medir factores como el neurodesarrollo en los niños, la cognición social y la regulación emocional en adultos, y la salud mental de todos los participantes para ayudar en el proceso de reintegración. En un proyecto piloto, financiado por fundaciones filantrópicas, han entrevistado a 200 civiles, además de los representantes de 150 excombatientes que viven en una zona tranquila a solo tres horas de distancia. "Necesitamos tener buenos datos para ver cómo está funcionando y para poder hacer intervenciones rápidamente si vemos que el proceso no está funcionando bien", dice Gómez.

Ha sido difícil para los investigadores – tanto aquí, como en otros lugares – estudiar si programas como estos pueden evitar que los combatientes vuelvan a la delincuencia, en gran parte porque a menudo es imposible rastrear los resultados de las personas que los atraviesan. "Simplemente asumimos que tiene un efecto, y no tenemos otra opción", dice Enzo Nussio, un experto en ciencias políticas del Instituto Federal Suizo de Tecnología, en Zúrich.

Sin embargo, Nussio y otros tienen esperanzas sobre los resultados en Colombia. El país tiene muchos más recursos para dedicar al esfuerzo que naciones como Burundi y Sudán, que han emprendido esfuerzos similares con poco éxito.

Los excombatientes, mientras tanto, se enfrentan a una mezcla de desafíos – algunos familiares y otros nuevos –. Al igual que los veteranos de otros conflictos, a muchos les resulta difícil estar cerca de personas que no entienden la experiencia del combate, dice Thomas Elbert, psicólogo de la Universidad de Konstanz, en Alemania. Pueden acercarse

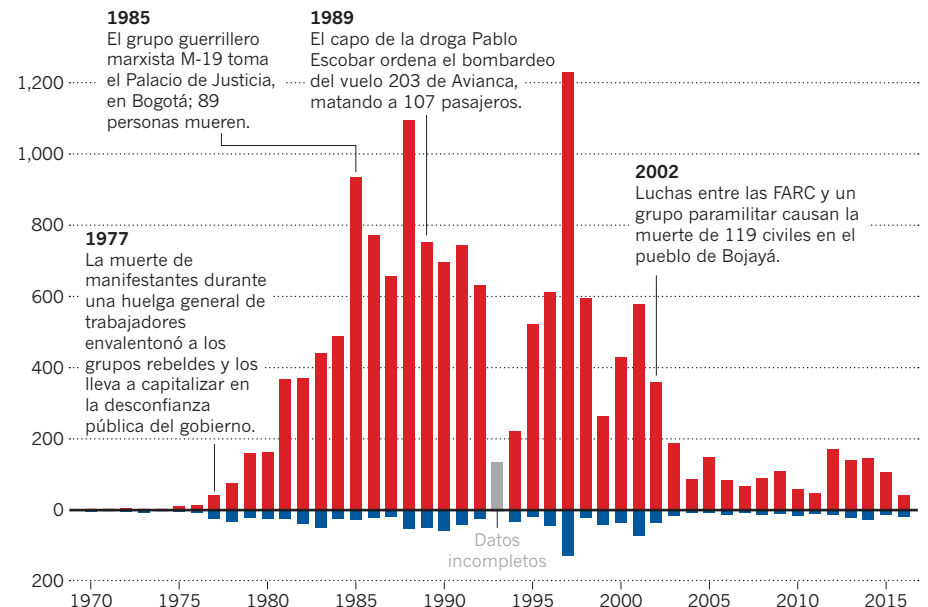
Un conflicto persistente

Un tratado de 2016 está ayudando a poner fin a décadas de violencia entre el gobierno colombiano y las Fuerzas Armadas Revolucionarias de Colombia (FARC), un grupo guerrillero de izquierda. Como resultado, miles de excombatientes de las FARC están ingresando a las denominadas 'zonas de tránsito' conocidas como espacios territoriales para el entrenamiento y la reincorporación (ETCR) que se encuentran en todo el país. En estas zonas, los excombatientes comenzarán un proceso para desmovilizarse y prepararse para el reingreso a la sociedad colombiana. Aún así, hay varios grupos guerrilleros que continúan luchando, incluidos 1.000 'disidentes' de las FARC que se han negado a abandonar sus armas. Ellos representan un peligro para los excombatientes y un desafío para los esfuerzos de establecimiento de la paz.

● ETCR ■ Actividad de los disidentes de FARC



■ Muertes ■ Secuestros



► a otros que han vivido en la violencia, que puede ser peligroso en un lugar como Colombia, donde los narcotraficantes y otros grupos armados todavía operan.

Colombia plantea algunos desafíos únicos, dice Gómez. A diferencia de aquellos que libran guerras civiles en muchos otros lugares, los guerrilleros colombianos no son impulsados por la raza o la religión, sino por ideología política. Reprogramar sus corazones y mentes podría requerir estrategias diferentes a las utilizadas para otros militantes radicalizados, criminales de guerra o asesinos en serie, y nadie sabe cuáles deberían ser esas estrategias.

Gustavo Tovar, presidente del consejo municipal de Vista Hermosa, teme que su pueblo –y el país– no estén listos para la ola de excombatientes. “Colombia está en el medio de esta metamorfosis”, dice. “Nosotros entramos en ella sin saber lo que estamos haciendo”.

EL PESO DEL PASADO

Al oeste de Meta se encuentra el Valle del Cauca, una región agrícola casi plana rodeada de montañas. Aquí, en un hotel lujoso con vistas a viñedos y plantaciones de caña de azúcar, el excomandante de las FARC Juan Carlos Sánchez está viendo videos en una computadora portátil. Diseñados para la investigación, los videos presentan diferentes tipos de altercados: una discusión entre dos personas, alguien que es golpeado por una silla, una persona que le mete una puñalada en la espalda a otro.

Sánchez se unió a la guerrilla en 1998 cuando vivía en Meta. No fue completamente por elección propia, dice. Cuando los militares colombianos abandonaron la región, las FARC se convirtieron en el gobierno de facto. La guerrilla persuadió a los locales para que se armaran en caso de que regresaran los paramilitares y el gobierno, diciéndoles que esta era la única forma de proteger a sus familias. Eventualmente, Sánchez se unió al grupo oficialmente.

Al principio, dice, los guerrilleros no atacaban a civiles –tan solo entrenaban a sus reclutas en política y tácticas de guerra–. Pero con el tiempo, las FARC se volvieron más violentas y desconfiadas de los demás. “Desde el comienzo de entrar en la organización, hasta el día que yo deserté, siempre me vine con preguntas”, dice, preguntas sobre él, sus camaradas y las personas que daban las órdenes, particularmente, su disposición a matar a antiguos aliados. Pero Sánchez calló sus dudas, porque cuestionar las órdenes lo llevaría a la muerte. En vez de eso, ascendió en las filas, liderando eventualmente a unos 25 combatientes.

Para 2005, se había vuelto más desilusionado con las FARC. Le entregaron una lista con nombres de informantes sospechosos a los que le ordenaron debía de matar; la lista incluía a niños de 12 y 13 años, así como a dos personas que él conocía desde su infancia. Él le ordenó a sus tropas que lo hicieran –algo que lo atormenta hasta el día de hoy–. “Uno es el que lleva la carga”, dice.

“Desde el comienzo de entrar en la organización, hasta el día que yo deserté, siempre me vine con preguntas”.

— JUAN CARLOS SÁNCHEZ



Luego, unos años después, los guerrilleros condenaron a su novia, también miembro de las FARC, por espionaje. Sánchez hizo planes para huir con ella, pero fue descubierto y tuvo que huir por su cuenta. Más tarde supo que ella había sido ejecutada.

Por tres años, Sánchez vivió escondido en Meta, viendo enemigos en todas partes. “Vivía en una zozobra permanente”, dice. Eventualmente, un amigo le contó sobre el programa de

rehabilitación del gobierno y se unió. Se mudó al Valle del Cauca y ahora trabaja en el campo para ganarse la vida. Él ha tenido consultas con un psicólogo y ha comenzado a leer la Biblia. “Entre los dos, me han ayudado a auto-perdonarme y a perdonar”, dice.

En el hotel, Sánchez mira los videos de altercados intercalados con una serie de preguntas: le preguntan si las imágenes le molestan, o si algunas de las partes involucradas debe ser castigada. Luego trata de identificar las emociones a través de una serie de caras.

“Estamos observando qué tipo de errores cometen”, dice Agustín Ibáñez, un neurocientífico de la Universidad Favaloro de Buenos Aires, Argentina, quien está administrando la prueba, una versión simulada de la misma prueba que ya le ha aplicado a unos 350 excombatientes colombianos. Ibáñez pretende discernir cómo el aislamiento de la sociedad y la exposición a la violencia pueden haber menoscabado su procesamiento emocional y su juicio moral¹. Los excombatientes –tanto lo de las FARC, como los paramilitares– tienen a tener problemas para distinguir entre las emociones, especialmente el miedo y la ira. Aunque no está claro si estos efectos persisten fuera del laboratorio, a Ibáñez y a su equipo les preocupa que los problemas emocionales puedan hacerles la vida aún más difícil a los excombatientes, como le ha ocurrido a excombatientes en otros lugares.

Él y Eduar Herrera, un psicólogo cognitivo de la Universidad Icesi, en Cali, comenzaron esta línea de investigación en 2014, trabajando con un grupo de paramilitares encarcelados por crímenes de guerra. Habían matado a un promedio de 33 personas cada uno, y algunos se les encontró responsables de masacrar a cientos.

“La primera vez, estábamos muy asustados”, recuerda Ibáñez. Los excombatientes no estaban esposados y se encontraron cara a cara con los investigadores. “Tienes la sensación de que podrían matar a todos, si así lo desearan”.

En 2017, los investigadores encontraron que una característica clave de los excombatientes es cómo juzgan la moralidad de una acción. La mayoría de los participantes condenarían un intento de envenenamiento, por ejemplo, incluso si no se lograra matar al objetivo. Pero el grupo de Ibáñez ha descubierto que es menos probable que los excombatientes condenen a alguien por un intento de asesinato fallido, razonando que si la víctima no murió, no hubo daño. Al mismo tiempo, es más probable que quieran castigar a las personas por daños que son claramente accidentales. Según su lógica, el resultado es más importante que la intención.

Basado en una pequeña muestra, parece que los paramilitares están más perjudicados en este sentido que los guerrilleros. Ibáñez dice que esta diferencia podría tener sentido: las personas que se unieron a los paramilitares por un salario podrían haber estado más atraídas hacia la violencia que aquellos que se unieron a

la guerrilla por razones ideológicas, aunque no hay forma de probar esto. Estos datos sugieren que los programas de rehabilitación no deberían tratar a todos los excombatientes de la misma manera.

Trujillo también ha encontrado diferencias notables en sus participantes. En un estudio de 624 exguerrilleros y exparamilitares², ella y sus colegas descubrieron que la capacidad de empatizar se dividía en 3 grupos: el 22% de los excombatientes funcionaba de forma muy similar a personas que no habían experimentado la violencia; el 32% tenía la capacidad de reconocer el dolor o la desgracia en los demás, pero no se veía tan afectado por ellos; y el resto no podía reconocer sentimientos como la angustia emocional en los demás, ni sentir empatía con ellos.

Los investigadores se preguntan si estos excombatientes –que se desmovilizaron voluntariamente– son similares a los 6.800 combatientes de las FARC que ingresaron al programa de reincorporación como parte del tratado de 2016, muchos de ellos por órdenes de sus comandantes. A diferencia de Sánchez, muchos siguen siendo decididamente ideológicos. Óscar Vega, por ejemplo, un excomandante de las FARC, delgado e intenso, que vive en la zona de tránsito cerca de Vista Hermosa, lleva a casi todas sus conversaciones al tema sobre las formas en que el gobierno y el sistema educativo de Colombia dañan a las personas. Él todavía vive por la causa. “Nuestros documentos y nuestra línea ideológica dice que por cualquier vía tenemos que llegar a la toma de poder, bien sea por la vía armada o bien sea por la vía política”, dice.

Trujillo está comparando varios tipos de terapias para determinar la mejor manera de ayudar a los excombatientes a mejorar su desempeño en las pruebas de empatía³. Ella y sus colegas están utilizando la electroencefalografía (EEG) para controlar la actividad cerebral de los excombatientes, con la esperanza de aprender cómo procesan la información⁴. En una investigación aún sin publicar, el equipo descubrió que los excombatientes son más rápidos en el reconocimiento de caras que los civiles –a pesar de que son más lentos en identificar las emociones que estas reflejan–. También son mejores en la realización de tareas de memoria que van acompañadas de imágenes violentas, como sangre o cadáveres. Las personas que los investigadores identificaron como víctimas de la violencia muestran el patrón opuesto: tales imágenes perturban su concentración y ralentizan sus respuestas. Los investigadores piensan que los circuitos neuronales de los excombatientes se han adaptado para reaccionar más rápido ante las amenazas.

El grupo de Trujillo asesora a la filial de Medellín de la ARN en su esfuerzo por rehabilitar a los excombatientes. Pero tratar de utilizar la ciencia para cambiar y hacer políticas puede ser como hacer esculturas con arena seca. “La investigación ha sido muy complicada porque es un tema muy nuevo,



“¿Cómo dejar ir esos pensamientos?”

— VIVIANA MISAS

no solo para Colombia, sino también para la literatura sobre neurociencia cognitiva y social”, dice. Los investigadores también se preocupan de que el gobierno pueda perder la paciencia. “Si no puedes encontrar algo sólido para demostrar lo que está sucediendo, no puedes proponer una solución”, dice José David López, un ingeniero que trabaja con Trujillo en la interpretación de los datos de los EEG, en la Universidad de Antioquia. “Lo necesitan ahora, no en diez años”.

UNA BATALLA DESDE ADETRÁS

En su muñeca, Viviana Misas lleva el nombre del bebé que perdió mientras vivía con el Ejército de Liberación Nacional (ELN), otro grupo guerrillero de izquierda que todavía está activo en Colombia. Misas se unió al ELN a los 15 años para alejarse de su familia, y llegó a amar la ideología y la camaradería que le proporcionaba. Pero luego, durante una marcha difícil se cayó y se lesionó, y en el proceso tuvo un aborto espontáneo. Sus compañeros la abandonaron y ella terminó en el hospital por un largo tiempo. Desconfiando de su lealtad después de eso, un compañero combatiente – su mejor amigo – la

traicionó y la delató al ejército colombiano. Ella fue capturada y aceptó desmovilizarse.

Al igual que Sánchez, no puede regresar a su casa en Medellín, por temor a que el ELN la mate como traidora. Aunque Misas disfrutaba de su trabajo como guía turística en el Valle del Cauca, la depresión le impide perseguir sus sueños. “Quisiera que cuando estoy triste ser una persona normal que solo se puso triste un momentito y no tiene pensamientos tan raros como los míos”, dice. Ella no ha visto ningún psicoterapeuta, pero su perro le brinda cierto consuelo, al igual que la religión. Aún así, sus pensamientos se vuelven oscuros; quizás, dice ella, la verdadera razón por la que se unió al grupo fue porque esperaba ser asesinada. “¿Cómo dejar (ir) esos pensamientos?”, se pregunta.

Según los datos de la ARN, más del 90% de los excombatientes en el programa tienen un problema psicosocial, como trastorno de estrés postraumático o ansiedad. El coordinador regional de ARN, Juan Fernando Vélez, dice que la salud mental es una de sus principales prioridades cuando trabaja con ellos. Los datos de Trujillo, dice, convencieron a su oficina de la necesidad de crear un programa especial de reintegración para personas con problemas psiquiátricos. “No podemos darle a la sociedad un individuo que no se siente bien consigo mismo”, dice.

Joshua Mitrotti, quien dirigió la ARN durante tres años antes de renunciar en marzo, dice que los programas de la agencia se basan en los esfuerzos llevados a cabo en América Central en la década de 1990, que proporcionaron formación vocacional y educación para los grupos armados. El apoyo psicosocial es un componente integral, dice.

El programa de la ARN para guerrilleros y paramilitares que se desmovilizaron voluntariamente incluyó 30 meses de servicios psicosociales, a cargo de unos 300 psicólogos y 65 trabajadores sociales, en promedio. Hasta el momento, 20.490 personas han completado el proceso de reintegración y la ARN dice que más del 70% se han reintegrado exitosamente a la sociedad.

Pero con decenas de miles de excombatientes en Colombia, simplemente no hay suficientes profesionales de la salud mental capacitados para brindar atención básica, y mucho menos terapia cognitiva intensiva. Por ello, algunos temen que el programa de reintegración pueda estar brindando un tratamiento deficiente. “No es que estén haciendo las cosas mal, pero están incompletas”, dice Herrera.

Uno de los desafíos es la dificultad de proporcionar tratamiento a adultos que no completaron la escuela primaria y no pueden leer, una habilidad requerida para algunas de las terapias habituales.

Mitrotti dice que la ARN ha estado ajustando los enfoques para hacerlos más apropiados. Según la ARN, el 30% de las personas que acudieron a los servicios psicosociales el año pasado lo hicieron sin ningún incentivo ▶



GREG KENDALL-BALL/NATURE

Con una economía lenta y una controversia continua sobre el acuerdo con las FARC, muchos en Colombia tienen sombrías esperanzas para el futuro.

► monetario (los excombatientes a menudo reciben un estipendio por participar en los programas). “No vienen porque se les paga, sino porque creen que necesitan el apoyo de nuestros profesionales”, dice Mitrotti.

Sin embargo, la ayuda para los miembros de las FARC que recién se han desmovilizado ha tardado en llegar. Andrés Restrepo, un sociólogo que trabaja en una zona de transición en Caquetá, dice que los excombatientes de las FARC no están recibiendo ningún tipo de atención de salud mental. Restrepo dice que la ARN les ha prometido la llegada de seis psicólogos a la región, pero incluso eso no sería suficiente para atender a los 1.000 excombatientes y a sus familias que viven allí ahora.

Restrepo teme que si estos individuos no se encuentran psicológicamente estables, el rechazo de la sociedad –incluidas sus propias familias– podría llevarlos nuevamente a la violencia. “Nadie les ayudó a imaginar una vida sin armas”, dice.

UN FUTURO INCIERTO

En Piñalito, un pequeño y polvoriento pueblo de casas de madera fluorescentes, en las afueras de Vista Hermosa, los civiles aún se están acostumbrando a la paz. “Es genial, no hay personas muertas”, dice Carlos García, un comerciante jubilado de edad avanzada. Recuerda escuchar de forma cotidiana los disparos, justo afuera de su puerta, cuando las FARC combatían a los paramilitares.

Ahora las calles son tranquilas y la gente visita los cafés. A algunos les faltan piernas – Colombia tiene una de las tasas más altas de víctimas de minas terrestres en el mundo, y Meta es una de las regiones más minadas-. Con el trasfondo de una ofensiva gubernamental contra el tráfico de cocaína, la caída de

“Nadie les ayudó a imaginar una vida sin armas”.

los precios del petróleo y el aumento del temor de que la paz sea solo temporal, la población de Piñalito parece tener pocas esperanzas para el futuro.

Y el acuerdo de paz en sí mismo está amenazado. Este mes, Colombia celebrará su elección presidencial, y el tema clave es si se debe renegociar el acuerdo para que sea menos favorable para las FARC. Muchos de los guerrilleros, mientras tanto, están perdiendo la fe en el proceso. Algunas de las zonas de tránsito a las que enviaron a los miembros de las FARC a vivir por un tiempo todavía no tienen agua corriente ni saneamiento. Y los programas de desarrollo agrícola y vocacional han tardado en arrancar. En todo el país, más de la mitad de

los guerrilleros han abandonado estas zonas, optando por jugarse la suerte en una sociedad que no es segura para ellos. Desde el acuerdo, cientos de exmiembros de las FARC han sido asesinados.

A medida que los excombatientes vuelven a la sociedad o se retiran a la jungla, los expertos se preocupan por el estigma que llevan – de estar afiliados a las FARC y el de las enfermedades mentales-. Matallana espera que una de las cosas que su investigación pueda hacer sea mostrar al público cómo el trauma afecta tanto a excombatientes como a civiles.

Los recursos son escasos y el problema es inimaginablemente complejo, dice Vélez. En última instancia, dice, el éxito de Colombia depende de la voluntad de su gente y de su capacidad para hacer las paces con el pasado. “No hay fórmulas mágicas”, dice. “Lo único que debemos entender es que todos necesitan –merecen– una segunda oportunidad”. ■

Sara Reardon trabaja como periodista para Nature desde Washington D.C. El Centro Pulitzer para el Reporteo de Crisis proporcionó el financiamiento del viaje para llevar a cabo este reportaje.

1. Baez, S. et al. *Nature Human Behav.* **1**, 0118 (2017).
2. Trujillo, S. P. et al. *J. Peace Psychol.* <https://doi.org/10.1037/pac0000255> (2017).
3. Trujillo, S. et al. *Front. Psychol.* **8**, 510 (2017).
4. Tobón, C. et al. *Soc. Neurosci.* **10**, 153–165 (2015).

COMMENT

ZOOLOGY Collaboration drove taxonomy pioneer **p.28**



HISTORY From tax to dice and astrology — what did Einstein really say? **p.30**

ARTIFICIAL INTELLIGENCE Studies of surveillance technology need ethical review **p.31**

SUSTAINABILITY Sea-bed mining code prompts concern for environment **p.31**

CLAUDIO BRESCEANI/AFP/GETTY



Chemist Youyou Tu, who discovered the malaria treatment artemisinin, was the first Chinese female scientist to win a Nobel prize.

Close the gender gap in Chinese science

Analysis shows that extending the age limit for grants boosts the number awarded to women, but more must be done to achieve parity, say **Ying Ma** and colleagues.

“**W**omen hold up half the sky” was a popular slogan in Mao Zedong’s China of the mid-twentieth century, intended to emphasize the equal importance of women in public and private life. But even though China used such slogans and had constitutional claims of gender equality decades

before many other nations, inequalities persist. By 2017, just 6% of the members of the Chinese Academy of Science were women.

In the 1980s and 1990s, advances in the country’s technological capacity generally involved importing knowledge. Now, China is focused explicitly on building its

own research and development (R&D) and innovation. Its R&D staff swelled from 3.2 million in 2009 to 5.8 million in 2016 (ref. 1), and the increased demand for talent has highlighted the need for more female scientists. Currently, women make up only about one-quarter of this workforce. At the same time, increased connections ►

► between China and the international community have made concerns about gender inequality more prominent.

Multiple governmental and scientific organizations in China have taken measures to promote women in science. Here we present the results of several initiatives undertaken by the National Natural Science Foundation of China (NSFC) (where X.G. and L.S. work; Y.H.Z. did so until March 2018).

Notably, after age limits for female grant applicants were extended, the percentage of women winning grants from a major fund for young scientists rose by 10% in one year. But there is still a long way to go.

MINDING THE GAP

Women in China have one of the highest rates of participation in the labour force when compared with women from both large developed and emerging economies, such as the United States, Germany, Brazil and India. This is a legacy of its planned-economy era, starting in 1949, when women's participation in the workforce was encouraged and protected. As late as 1988, women made up 48% of the labour force in China, and women's average earnings were 84% of those for men. By 2002, however, 10 years after the country moved to a market economy, women made up 46% of the labour force and their earnings were 79% of those of men².

Many universities in China have adopted a policy of 'promote or leave'. This means that scientists gain a permanent position only if they pass an evaluation at the end of a 6-year probationary period, which often coincides with women's child-bearing years.

Similar to other countries, China has a 'leaky pipeline' for women in science — fewer women advance through each stage of a scientific career. In 2016, 53% of master's students and 39% of doctoral students in China were women³. That proportion falls to 14% for recipients of the NSFC's Distinguished Young Scholars Award, which helps rising researchers under 45 to become leaders in their fields.

In 2010, a joint document from the Central Committee of the Communist Party of China and the State Council called for the creation of policies to help talented men and women balance work and family. It advocated for a more equitable gender ratio in professional workplaces. In 2011, the Ministry of Science and Technology and the National Women's Federation jointly issued a policy document to champion the development of women in science and technology careers.

In 2010, a survey of the NSFC's applicants found that about 70% of women and 24% of men supported a policy that sought to redress historical disadvantages through affirmative action. Measures around maternity and parental rights were

GENDERED ATTITUDES

Nearly 6,000 scientists across China were asked whether they agree with the following statements. The survey found viewpoints, especially among men, that could hold back women's careers.

"Women are 'not suitable' for research work."



"A man's success is measured by his career, while a woman's success is measured by her family."



"Men make better project leaders."



supported by majorities of both genders⁴.

Most grant and job applications in China already restrict eligibility by age, so changing these requirements offered a way to support female scientists. Thus, in 2011, the NSFC increased the age limit for women applying to its Young Scientist Fund from 35 to 40, while that for men remained at 35 (one of us, X.G., was involved in making this decision). This programme is the second-largest of the NSFC funds and the main way in which early-career scholars in basic science receive national funding. As of 2016, the programme represented 13.8% of the roughly US\$4.1-billion budget the NSFC spent on projects, and financed 39% of all individual projects.

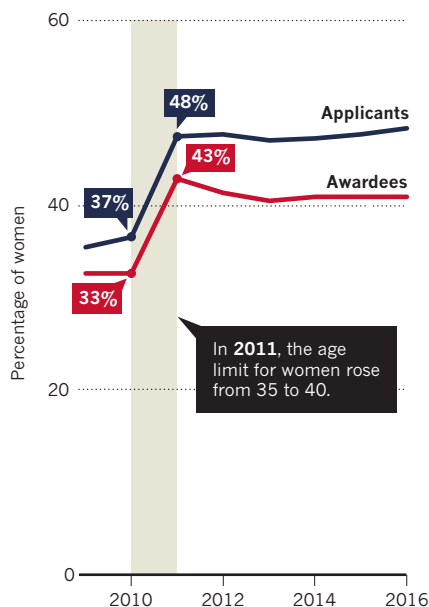
Higher age limits (38 for men, 40 for women) were also established for a new

programme, the Excellent Young Scientist Fund. This supports about 400 projects a year and represents 2.2% of the NSFC budget. Another new policy allowed women to apply to extend NSFC project terms (but not funding amounts) by up to 24 months for maternity leave.

Also in 2011, the NSFC pledged to increase the number of female scientists on review panels, although it did not set a quota. It invited review panels to consider prioritizing female applicants when all else was equal; enhanced the publicity surrounding research findings by female scientists financed by its programmes; and started to collect statistical data about the gender of applicants and awardees⁵. In 2016, the Chinese Academy of Science and Technology for Development (CASTED) surveyed more than 5,800 scientists about their attitudes to gender roles and recent policies (an effort led by Y.D.Z. and Y.M.).

MORE GRANTS FOR WOMEN

Policies to accommodate parenthood increased female applicants and awardees for a Chinese fund for young scientists.



EFFECTS AND EXPECTATIONS

What happened? Raising the age bar in 2011 saw the percentage of women applying for the Young Scientist Fund increase from 37% to 48% (see 'More grants for women'). Applications from women soared by 94% to 25,694; about one-third were aged 36–40. Applications from men went up by 23%, to 28,397 in the same period. The year before, applications rose by 25% for men and 31% for women⁴, partly owing to swelling numbers of people gaining science and engineering PhDs: a 58% increase from 2006 to 2016.

The percentage of female award recipients jumped from 33% to 43% in 2011, and has remained at about this level. Despite this increase, a female scientist's chance of winning one of these grants has declined slightly, from 21% in 2010 (compared with 24% for men) to 19% in 2016 (26% for men). A lower success rate for women has been found in other programmes in

China and internationally⁶. It is hard to pin this on discrimination, differences in application styles or other reasons. We did observe that the success rate of women applicants aged 36–40 is lower than that of younger women. Despite this, more women are now receiving these awards.

Women's representation on review panels went up by 45% between 2010 and 2017, to 13.3%. That is still low, but consistently higher than rates seen from 1986 to 2009, which fluctuated around 6%.

No women have yet applied to extend their project terms for pregnancy or child rearing. However, we think that many would have applied for extensions had they known about the policy: in the 2016 survey, 60% of female scientists indicated that they had never heard of it.

Although these measures have had relatively little time to influence the scientific enterprise in China, more than 70% of female scientists polled expect that each policy will have a positive effect in their discipline.

Men who responded to the poll are less enthusiastic. About 60% thought that raising age limits for female applicants would have a positive effect on their field, as did 53% for extending project terms for maternity leave. Only 39% of men thought that increasing the number of women on review panels or favouring female applicants when all else was equal would be good for their fields.

WIDESPREAD PROBLEM

Discrimination and bias towards women in the workplace in China, as elsewhere, is all too common. A 2015 survey conducted in Beijing found that 87% of female university students encountered gender discrimination in their job hunt.

Even among scientists, the CASTED survey found bias and burdens that must affect women's careers (see 'Gendered attitudes'). More than 20% of men and around 10% of women agreed with the statements "A man's success is measured by his career, while a woman's success is measured by her family" and "men make better project leaders". For the second statement, 48% of men and 81% of women disagreed. (We did not ask inverted versions of the questions, such as whether women make better project leaders.)

Women feel the effects of these attitudes. Thirty-two per cent of female scientists reported that they encountered employers in their first job search who wanted to recruit only men. Given that 84% of the women surveyed were aged 45 or under, we must assume that most of this pool had experienced

"Eighteen per cent of male scientists say that their spouse does most of the housework."



Chemist Youyou Tu worked with pharmacologist Lou Zhicen (left) on traditional Chinese medicine in the 1950s, when women's participation in the workforce in China was encouraged and protected.

discrimination in recent decades.

Unequal responsibilities for child rearing, care for older people and other domestic labour also hinder women's career advancement in China, as has been reported for the United States⁷. Among married scientists in our survey, 30% of women compared with just 6% of men reported doing most housework themselves. And 2% of female researchers and 18% of male scientists say that their spouse does most of the housework. What's more, the gradual lifting of China's one-child policy from 2013 has placed more parental responsibilities on women.

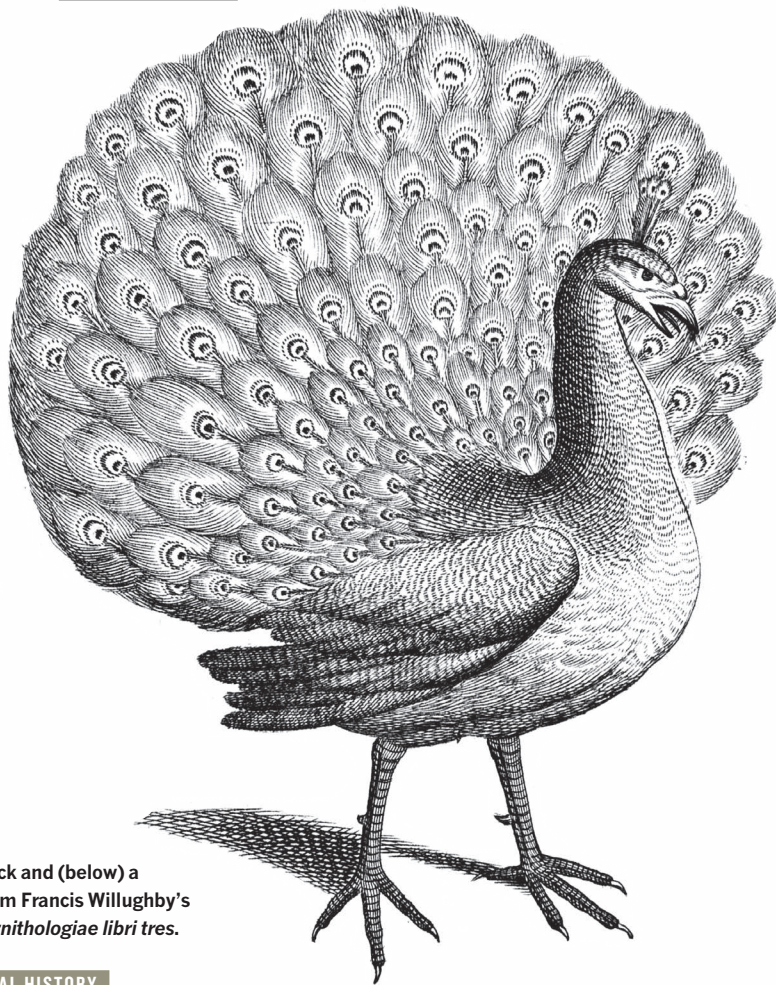
Women are less likely than men to change location to advance their career. One scholar explored why only 11.4% of Chinese recipients of funds from a German programme for visiting researchers in 2011 were female, and concluded that the 'price for mobility' was much higher for women than for men, because of marriage and family⁸.

Chinese society in general and the scientific community in particular are undergoing big transformations. The optimistic view from our perspective is that straightforward policy changes are helping. However, as a funding agency, the NSFC's role is limited. It is up to institutions to make decisions in hiring, appraising and promotion. The next step would be for the rest of China's research system to explicitly acknowledge

that various barriers in science prevent women from enjoying a level playing field with men, and to take measures to eliminate the existing gender bias. ■

Ying Ma and Yandong Zhao are senior researchers at the Institute of Science, Technology and Society, Chinese Academy of Science and Technology for Development (CASTED), Beijing, China. **Xu Gong** is a senior researcher and **Li Sun** is a researcher at the Bureau of Policy, National Natural Science Foundation of China, Beijing, China. **Yonghe Zheng** is a professor in the Faculty of Education at Beijing Normal University, Beijing, China.
e-mails: may@casted.org.cn; zhaoyd@casted.org.cn

1. National Bureau of Statistics and Ministry of Science and Technology. *China Statistical Yearbook on Science and Technology 2017* (China Statistics Press, 2017).
2. Li, C. & Li, S. *Sociol. Stud.* **2**, 94–117 (2008).
3. Liu, C. (ed.) *Educational Statistics Yearbook of China 2016* (China Statistics Press, 2017).
4. Zhao, Y., Ma, Y. & Liao, M. *Bull. Natl Nat. Sci. Found. China* **5**, 403–409 (2016).
5. Gao, R. *Bull. Natl Nat. Sci. Found. China* **2**, 94–96 (2011).
6. Bornmann, L., Mutz, R. & Daniel, H.-D. *J. Informetrics* **1**, 226–238 (2007).
7. Schiebinger, L. & Gilmartin, S. K. *Academe* **96**, 39–44 (2010).
8. Leung, M. W. H. *Sustainability* **5**, 2644–2663 (2013).



A peacock and (below) a dodo from Francis Willughby's 1676 *Ornithologiae libri tres*.

NATURAL HISTORY

A prodigious namer of nature

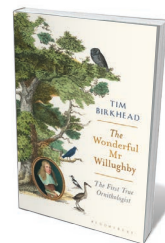
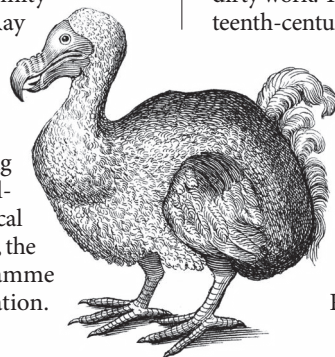
Elizabeth Yale relishes a biography of a seventeenth-century polymath with a notable gift for collaboration.

Francis Willughby was a wide-ranging virtuoso in a virtuosic age. Remembered for a pioneering study of bird classification, the seventeenth-century natural historian pursued interests in entomology, botany, linguistics, games and chance, and the reform of biological classification. That he is not better known may be put down to his death at 36. In *The Wonderful Mr Willughby*, ornithologist Tim Birkhead brings his creative energies and contributions to life.

Born in England in 1635, the only son in an established family of the gentry, Willughby inherited estates in Warwickshire and Nottinghamshire. At the University of Cambridge, where polite young men usually acquired a smattering of culture and influential connections, he took a different path — into scientific discovery. He dived into the

“new sciences”, reading the works of Galileo Galilei, Francis Bacon and René Descartes. And he took copious notes, organized by topic, in his commonplace book — the database of the era.

A sociable man, Willughby found friends who spurred him on. As Birkhead relates, the most important was Trinity College fellow John Ray. Ray encouraged Willughby's interests in mathematics and took him botanizing. It was during these jaunts that Willughby observed puzzling transformations in caterpillars that sparked entomological discoveries. In the late 1650s, the pair embarked on a programme of “chymical” experimentation.



The Wonderful Mr Willughby: The First True Ornithologist
TIM BIRKHEAD
Bloomsbury (2018)

“Chymistry”, as practised by Robert Boyle and other natural philosophers, was then evolving from medieval alchemy to modern chemistry. It sought the transmutation of base metals into gold even as it was harnessed for applications such as weapons manufacture. Eventually, Willughby and Ray criss-crossed England and Wales on

their birding and plant-hunting expeditions.

Around 1662, they set themselves an ambitious goal: observing, describing and classifying all species. They felt that both the literature and the nomenclature sowed confusion. Swiss naturalist Conrad Gessner's *History of Animals* (1551–58), for instance, mixed ancient knowledge with observation. By contrast, Ray and Willughby grounded their system in precise anatomical description, distinguishing between even closely related species. Beginning with British species and extending to mainland Europe, they established a taxonomy that would be built on by centuries of naturalists, including Carl Linnaeus in the mid-eighteenth century. Dividing birds into land and water fowl, they deployed attributes such as beak shape to create a branching classification key.

Willughby thrived on collaboration, and used his wealth to enable it. In 1662, Ray resigned his college fellowship, rather than subscribe to the Act of Uniformity passed by Parliament to fortify Charles II's newly restored monarchy. Willughby invited his mentor into his household. The next year, Willughby was elected an “original fellow” of the Royal Society, and he and Ray, with Ray's students Philip Skippon and Nathaniel Bacon, ventured on a tour across Europe.

They attended university lectures and visited cabinets of curiosity — troves of exotica where they handled a hornbill's head and an elephant's tail. They collected birds' eggs and a book of paintings of birds and fish from Leonard Baldner, keeper of forests in Strasbourg, now part of France. In rented rooms, they dissected and drew fish from Venice markets, a servant often doing the dirty work. They visited the museum of sixteenth-century naturalist Ulisse Aldrovandi

in Bologna and attended human dissections. Of this very Protestant crew, Willughby alone braved the dusty roads of Catholic Spain, which he viewed as a forbidding closed society.

After they returned to England in the mid-1660s, Ray stayed on at Willughby's

estate as the latter married, had children and managed his lands. Birkhead gives a wonderful sense of the pair's delight in nature, even as Willughby, never robust, began to have recurring fevers. Inspired by physician William Harvey's discovery of blood circulation, published in 1628, Willughby contemplated the movement of sap in trees years before the subject surfaced in the Royal Society's journal, *Philosophical Transactions*. He was the first to classify insects by their metamorphoses, recognizing that a caterpillar, pupa and butterfly were life stages of one insect, not separate species. He asked astute questions, such as which birds survive winters by migrating. He observed the life cycle of a leaf-cutter bee, later named after him — *Megachile willughbiella*. He even wrote a study on games, from football to cards.

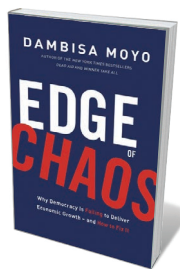
Birkhead's account is vividly textured, drawing from his collaborations with science historians. We follow Willughby from seabird nesting grounds on the Isle of Man to glass-making factories in Murano, Venice. Willughby's letters and notebooks, full of his swift, impatient writing, tell how avidly he worked. The strangeness of his scientifically liminal century shines through, exemplified by an "insect" collected in Italy, a fake made from a moray eel's jaws and a thorny plant. Birkhead tightens the links between Willughby's work and modern biology, confirming that he and Ray identified some 90% of around 200 bird species often seen in England and Wales.

As Birkhead emphasizes, the bond between the restless Willughby and the more restrained Ray was extraordinarily fruitful. Yet there were challenges, not least differences in social circumstances. Willughby was a gentleman, Ray a blacksmith's son — disparities they finessed in life. That became more difficult after Willughby's death. In exchange for an annuity, the family expected Ray to educate Willughby's children; he was reluctant. They also resented Ray's control over Willughby's posthumous legacy. They quarrelled over access to Willughby's collections and papers as Ray produced *The Ornithology* (1676), *The History of Fishes* (1686) and *The History of Insects* (1710), based on his joint work with his friend. Subsequently, historians have struggled to divide the credit, sometimes favouring one man, sometimes the other.

"This game of spot-the-genius is inappropriate and unhelpful," writes Birkhead. He invites us to see a scientific life well lived, rich with ideas, adventure and companionship — and, in Willughby's profound collaboration with Ray, two very different personalities who saw further because they worked together. ■

Elizabeth Yale is a lecturer in the Department of History at the University of Iowa. She is the author, most recently, of *Sociable Knowledge: Natural History and the Nation in Early Modern Britain*. e-mail: elizabeth-yale@uiowa.edu

Books in brief



Edge of Chaos

Dambisa Moyo LITTLE, BROWN (2018)

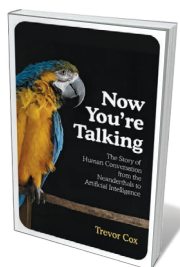
Does the "new normal" in many democracies, from high unemployment to political turmoil, make them poor models for sustainable growth? In this trenchant analysis, economist Dambisa Moyo explores that provocative question. She examines growth across the political spectrum, from China to the United States, and probes entangled challenges such as debt and protectionism. Unsurprisingly, she points to an urgent need for political reform. Her blueprint for that (including civics courses for the electorate) is ambitious, but, as she asserts, "All the easy choices are behind us".



Eye of the Shoal

Helen Scales BLOOMSBURY SIGMA (2018)

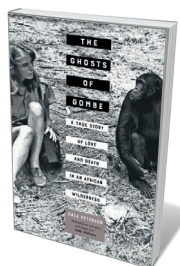
Marine biologist Helen Scales's *Spirals in Time* (2015) opened up a whorled wonderland of marine molluscs. This gifted writer now deep-dives into piscine realms. Scales, whose research has spanned the South China Sea and Australia's Ningaloo coral reef, weaves the history of ichthyology with explorations of adaptations, such as how glycoproteins act like 'antifreeze' in the blood, and why shoaling saves energy. Perhaps most beguiling are the hums and pops of fish 'calls', which the creatures sense through the lateral line — a series of organs that effectively turn their bodies into giant ears.



Now You're Talking

Trevor Cox BODLEY HEAD (2018)

On average, humans utter 500 million words over a lifetime. And it's a crazily complex process, as acoustic engineer Trevor Cox reveals in this intensive survey. Speaking involves "anatomical gymnastics" linked to multiple brain regions; hearing is a subtle decoding of tone, timbre and sense. Cox's investigation sweeps from the putative protolanguage of human ancestor *Homo heidelbergensis* to the likelihood of creative algorithmic discourse. In between, he looks at the infant's acquisition of language, the neuroscience of beatboxing (vocally mimicking percussion instruments) and much more.



The Ghosts of Gombe

Dale Peterson UNIVERSITY OF CALIFORNIA PRESS (2018)

In July 1969, Ruth Davis — a volunteer at Jane Goodall's chimpanzee research centre in Gombe, Tanzania — disappeared. Her body was found below a waterfall six days later. Goodall biographer Dale Peterson probes the tragedy and its convoluted context in forensic detail, casting back and forth from the centre's primatological findings to the human stories of its researchers. Peterson's engrossing, sometimes dizzyingly kaleidoscopic narrative is bookended by nuanced analyses of how Davis might have died, and the aftershocks that still rock those who knew her best.



On Color

David Scott Kastan and Stephen Farthing YALE UNIVERSITY PRESS (2018)

Artistic innovator Paul Cézanne accurately noted that colour is a collaboration between mind and world. So remind literary scholar David Scott Kastan and artist Stephen Farthing in this vivid and erudite tour of a phenomenon that entwines microphysics and electromagnetics with human physiology and cognition. Their march through ten hues drives home why much of culture is deep-dyed in colour, from political affiliations (think the Greens, or Ireland's Orange Order) to blue notes in music, "uncannily microtonal slides and bends" expressive of emotional subtleties. **Barbara Kiser**

PHYSICS

Einstein said that — didn't he?

As the physicist's papers reach volume 15, **Andrew Robinson** sifts attributed quotes.

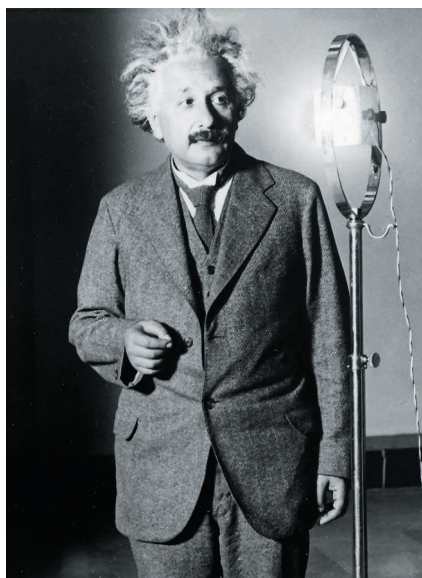
Beyond his towering contribution to physics, Albert Einstein was an avid commentator on education, marriage, money, the nature of genius, music-making, politics and more. His insights were legion, as we are reminded by this month's publication of volume 15 in *The Collected Papers of Albert Einstein*. Even the website of the US Internal Revenue Service enshrines his words (as quoted by his accountant): "The hardest thing in the world to understand is the income tax."

"There appears to be a bottomless pit of quotable gems to be mined from Einstein's enormous archives," notes Alice Calaprice, editor of *The Ultimate Quotable Einstein* (2011); one detects a hint of despair. Indeed, Einstein might be the most quoted scientist in history. The website Wikiquote has many more entries for him than for Aristotle, Galileo Galilei, Isaac Newton, Charles Darwin or Stephen Hawking, and even than Einstein's opinionated contemporaries Winston Churchill and George Bernard Shaw.

But how much of this superabundance actually emanated from the physicist? Take this: "Astrology is a science in itself and contains an illuminating body of knowledge. It taught me many things and I am greatly indebted to it." These lines, displayed by some astrology websites as Einstein's, were exposed as an obvious hoax by the magazine *Skeptical Inquirer* in 2007. The real source was the foreword to a reissued book, *Manuel d'astrologie* (1965), first published by Swiss-Canadian astrologer Werner Hirsig in 1950. Einstein's only known comment on astrology is in a 1943 letter to one Eugene Simon:

I fully agree with you concerning the pseudo-science of astrology. The interesting point is that this kind of superstition is so tenacious that it could persist through so many centuries.

Among the hundreds of quotes that Calaprice notes are misattributed to Einstein are many that are subtly debatable. Some are edited or paraphrased to sharpen or neaten the original. "Everything should be made as simple as possible, but no simpler" might, says Calaprice, be a compressed version of lines from a 1933 lecture by Einstein: "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience." More certain is the provenance of "The most incomprehensible thing about the Universe is that it is comprehensible." That rewords a



Albert Einstein in Caputh, Germany, in 1929.

passage in a 1936 article in the *Journal of the Franklin Institute*: "The eternal mystery of the world is its comprehensibility... The fact that it is comprehensible is a miracle."

Even "God does not play dice", arguably Einstein's most famous quote, isn't quite his words. It derives from a letter written in German in December 1926 to his friend and sparring partner, theoretical physicist Max Born. It is published in the new volume of Einstein's papers, in which the editors comment on its "varying translations" since the 1920s. Theirs is: "Quantum mechanics... delivers much, but does not really bring us any closer to the secret of the Old One. I, at any rate, am convinced that *He* does not play dice." Einstein does not use the word 'God' (*Gott*) here, but 'the Old One' (*Der Alte*). This signifies a "personification of nature", notes physicist and Nobel laureate Leon Lederman (author of *The God Particle*, 1993).

Einstein's name has also been affixed since his death to quotes from elsewhere. "The definition of insanity is doing the same thing over and over and expecting different results," for instance, was traced by Einstein archivist Barbara Wolff to US writer Rita Mae Brown's *Sudden Death* (1983). "Not everything that can be counted counts, and not everything that counts can be counted," was penned by sociologist William Bruce Cameron in his *Informal Sociology* (1963).

This cosmos of quotes — real, massaged and faked — speaks to Einstein's status. More than 60 years after his death, his fame remains paramount. I feel there are at least four

reasons why we are still fascinated by him.

One is that Einstein's discoveries are elemental and existential, unifying concepts of space and time, mass and energy and forces. They shifted our picture of reality. And he made more than a stab at explaining them to the non-physicist. Hence his part-joking encapsulation of relativity to the hungry press in 1921, on his first visit to the United States: "It was formerly believed that if all material things disappeared out of the universe, time and space would be left. According to relativity theory, however, time and space disappear together with the things."

There is also widespread empathy for Einstein's resilience in his long struggle for security. His performance at his German school was good, but far from brilliant; he disliked the school for its regimentation and eventually abandoned it. He failed to get an academic position after graduation from university, partly because he mocked his physics teachers. In 1901, although semi-starving, he recognized the value of not conforming. He wrote to his fiancée that "impudence" was his "guardian angel". It would guide him throughout his life.

Einstein was also highly engaged politically and socially, and often in the public eye. He supported the creation of a Jewish home in Palestine, helped to establish the Hebrew University of Jerusalem, and in 1952 was offered Israel's presidency. Yet he had written in a speech in 1938: "My awareness of the essential nature of Judaism resists the idea of a Jewish state with borders, an army, and a measure of temporal power." In 1933, he had publicly opposed Nazi Germany, fleeing to the United States by way of Britain, under some risk of assassination. Despite encouraging US president Franklin D. Roosevelt to build an atomic bomb in 1939, he was horrified by its use in 1945 in Japan. He spoke out against racial and ethnic discrimination in the United States. In the 1950s, he trenchantly criticized the hydrogen bomb and McCarthyism, and, right up to his death in 1955, he was targeted for deportation as a Soviet agent by FBI director J. Edgar Hoover.

Finally, there is Einstein's ineffable wit. It is encapsulated by this aphorism, composed for a friend in 1930 (really: I've checked with the Einstein Archives in Jerusalem): "To punish me for my contempt of authority, Fate has made me an authority myself." ■

Andrew Robinson is the author of *Einstein: A Hundred Years of Relativity*.
e-mail: andrew@andrew-robinson.org

ULLSTEN BILD VIA GETTY

Correspondence

Ethics review for AI surveillance studies

Scientists who develop algorithms based on user data face a moral dilemma: if their work is subverted to manipulate democracy or to support oppressive regimes, they could become part of something they would not knowingly endorse.

In common with other fields, research on artificial intelligence (AI) and machine learning needs to be subject to approval from institutional review boards and compliance with data protection. In my view, journals should demand this as a condition of publication. Data scientists in industry must also adhere to professional guidelines from organizations such as the IEEE (see go.nature.com/2vt6ngr). And research that combines academic and corporate interests should be disclosed, as in other fields.

As surveillance is combined with intelligent forms of behaviour-change technology, a new social contract around data is needed (see also H. Shah *Nature* **556**, 7; 2018). Unlike the media companies, governments and organizations that use surveillance data, those who are monitored do not benefit. Nor do they have any control over how their data are used (uninformed consent), as was poignantly illustrated by the Cambridge Analytica scandal (see *Nature* **555**, 559–560; 2018).

Rafael A. Calvo, Dorian Peters
University of Sydney, Australia.
rafael.calvo@sydney.edu.au

Poor meta-analyses pollute the literature

Meta-analysis of published data is important in evidence-based medicine. However, it is an experiment-free route to rapid publication and so is open to abuse. Extra vigilance by peer reviewers and journal editors is called for to prevent redundant and conflicted meta-analyses from corrupting the literature.

China produced 63% of meta-analyses of genetic associations

in 2014, and most of those results are misleading (J. P. A. Ioannidis *Milbank Q.* **94**, 485–514; 2016). Pressure to publish may be responsible, given that doing actual experiments takes much longer and can yield insufficient clinical data. And skilful presentation is often all it takes to disguise a poor-quality meta-analysis.

Conclusions from arbitrarily merging results of variable quality will not resolve problems and should not guide clinical practice. A rigorous meta-analysis requires meticulous evaluation of the literature. And even high-quality meta-analyses in leading journals still need constant clinical testing to ensure that current guidelines for treatment remain valid.

Yong Fan* *Hong-Hui Hospital, Medical College of Xi'an Jiaotong University, Xi'an, China.*
wqnsipine@163.com

**On behalf of 4 correspondents (see go.nature.com/2hxcccj for full list).*

Count the costs of sea-bed mining

The International Seabed Authority (ISA) is negotiating a mining code to allow commercial deep-sea mining of minerals to start worldwide. At Greenpeace, we argue that we should instead be developing a sustainable circular economy that reduces the use of virgin materials.

There is a huge demand for minerals in the computing, renewable-energy and mobility sectors. So far, the ISA has approved 29 exploration contracts in the Pacific, Indian and Atlantic oceans. Next year, the Canadian company Nautilus Minerals plans to mine copper, zinc and gold at depths of 1,500–2,000 metres in waters off Papua New Guinea.

In our view, the ISA should take more account of the biological and ecological impact of these mining activities. It conspicuously lacks an environmental committee, for example. Proper oversight is crucial, because sea-bed mining risks wiping out pristine habitat and potentially

unknown species.

Mining-induced loss of biodiversity in the deep sea is likely to last forever on human timescales, given the slow natural rates of recovery in affected ecosystems (C. L. van Dover *et al. Nature Geosci.* **10**, 464–465; 2017). We should instead be recycling the valuable materials contained in the 90% or so of the world's electronic waste that is currently illegally traded or dumped (see also go.nature.com/2toh2vr).

Sebastian Losada *Greenpeace International, A Coruña, Spain.*
Pierre Terras *Greenpeace International, Ar Bonoù, France.*
pierre.terras@greenpeace.org

Use SDGs to guide climate action

The United Nations Agenda for Sustainable Development commits all countries to attaining 17 goals (SDGs) and 169 targets by 2030, including SDG13's action to combat climate change and its impacts (go.nature.com/2r1wf72). Notwithstanding this goal's long-term benefits and synergies across other SDGs, climate action could have trade-offs with several of the SDG targets (see also M. Nilsson *et al. Nature* **534**, 320–322; 2016). We suggest that the SDGs should be used as reference points to map relationships between climate action and sustainable development.

For example, climate-mitigation policies in carbon-intensive and energy-exporting countries could slow economic growth (counter to target 8.1) or impair industrialization (target 9.2) in some sectors while boosting others. For end uses of energy alone, an estimated US\$3.5 trillion needs to be invested annually from 2016 to 2050 to adhere to a warming trajectory well below 2°C (go.nature.com/2jpmtbs).

Climate policies can also be socially and economically regressive, exacerbating inequality and poverty (targets 1.1 and 1.2) through impacts on land and food prices (target 1.4)

and putting smallholders at risk (target 2.3). And some national climate-adaptation programmes have been linked with violent conflict (B. K. Sovacool *World Dev.* **102**, 183–194; 2018).

Effective policy on climate action and sustainable development requires researchers and decision-makers to be mindful of such trade-offs and of how they could risk undermining the social and political support needed for climate action.

Francesco Fusco Nerini* *KTH Royal Institute of Technology, Stockholm, Sweden.*
francesco.fusonerini@energy.kth.se
**On behalf of 10 correspondents (see go.nature.com/2khyt96 for full list).*

Blobel's Nobel — why so slow?

Goran Hansson's assurance that Nobel prizes continue to recognize the potential impact of a discovery is undermined by the lengthy interval before the award is won (*Nature* **556**, 31; 2018). Cell biologist Günter Blobel is a case in point (see S. Simon *Nature* **556**, 32; 2018).

I attended a research seminar by Blobel in the late 1970s as an undergraduate biochemist. Three years or so after Blobel published the papers Simon mentions, our lecturers all acknowledged the central importance of that work. The Nobel committee only caught up 20 years later, finally recognizing what was by then in undergraduate textbooks, when Blobel was in his 60s.

This interval is typical (see S. Fortunato *et al. Nature* **508**, 186; 2014), with prize recipients often past retirement age — hardly a reward for emerging excellence.

Academic recruitment committees, too, tend to favour proven success over promising talent. Today, the postdoc who faced failure after failure in validating a brilliant intuition would not be rewarded with continued support as Blobel was. **William Bains Rufus Scientific, Melbourne, Royston, UK.**
william@rufus-scientific.com

Faulty replication can sting

Inappropriate cellular inflammation can cause disease. It emerges that the protein SAMHD1 prevents the release of newly replicated DNA from the nucleus, blocking an undesirable pro-inflammatory response. [SEE ARTICLE P.57](#)

MADZIA P. CROSSLEY
& KARLENE A. CIMPRICH

Cells need to distinguish between their own DNA and that of viruses. To solve this problem, plant, fungal and animal cells store their DNA in the nucleus, and respond to DNA in the cytoplasm by activating an inflammatory response. However, under certain abnormal circumstances, host-cell DNA can also accumulate in the cytoplasm, triggering inappropriate inflammation. On page 57, Coquel *et al.*¹ define a role for the protein SAMHD1 in preventing this cytoplasmic host-DNA build-up. The authors' findings have implications for immune disease and cancer.

SAMHD1 is a nuclear protein that chemically inactivates nucleotides², preventing them from being used to build DNA. Inherited SAMHD1 mutations cause the rare inflammatory disorder Aicardi-Goutières syndrome, which involves increased production of proteins called interferons that activate the immune system. Under normal circumstances, interferons are produced only in response to infection, and help to fight off viruses³.

Aicardi-Goutières syndrome can also be caused by mutations in the enzymes RNase H2 and TREX1, which degrade specific nucleic acids³. In TREX1-mutant cells, single-stranded DNA (ssDNA) fragments accumulate in the cytoplasm and activate a DNA-sensing pathway, dubbed cGAS-STING, triggering interferon production³. Before the current study, the cause of chronic inflammation in cells harbouring SAMHD1 mutations was unknown. Coquel *et al.* found that, as in TREX1-mutant cells, ssDNA accumulates in the cytoplasm of SAMHD1-deficient human cells. This leads to interferon production, mediated by the cGAS-STING pathway.

Why does SAMHD1 deficiency cause this defect? Dynamic structures called replication forks form at sites where double-stranded DNA is unwound, enabling each strand to be duplicated during DNA replication. The progression of forks along DNA can slow or stall if unusual DNA structures or damaged DNA block their paths⁴, or if nucleotide levels become depleted (Fig. 1a). Several proteins then cooperate to circumvent or repair the damaged site and restart replication. In some cases, degradation of newly synthesized DNA

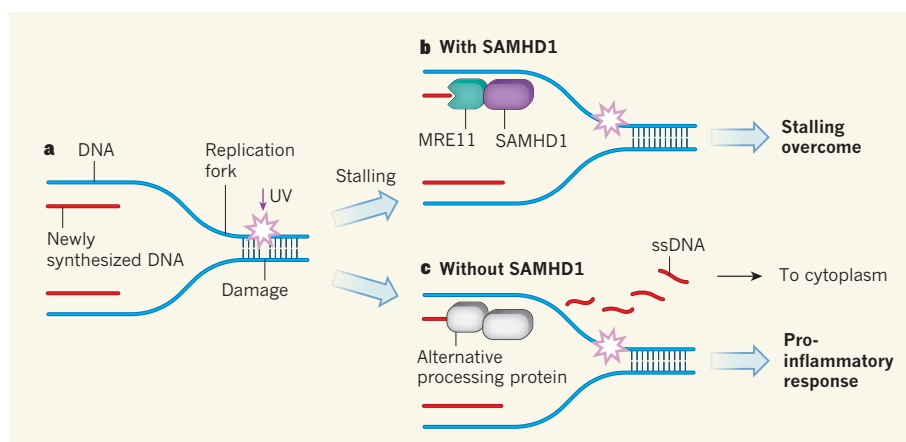


Figure 1 | SAMHD1 at DNA-replication forks. **a**, During DNA replication, DNA is unwound and duplicated to produce newly synthesized DNA at dynamic structures called replication forks. Forks can stall during replication if they hit unusual DNA structures (not shown) or DNA that has been damaged, for example by ultraviolet (UV) light. **b**, Coquel *et al.*¹ report that the protein SAMHD1 directs the nuclease enzyme MRE11 to degrade newly synthesized DNA — a process called fork resection, which is crucial for overcoming stalling. **c**, In SAMHD1-deficient cells, MRE11 activity is lacking, and the newly synthesized DNA at stalled forks is processed by alternative proteins, resulting in the release of single-stranded DNA (ssDNA) into the cytoplasm. The ssDNA fragments accumulate and trigger pro-inflammatory responses.

at stalled forks (a process called fork resection) promotes DNA remodelling or activation of repair pathways that help to restart DNA replication. When the authors artificially induced replication-fork stalling in SAMHD1-deficient cells, cytoplasmic ssDNA levels and interferon responses increased. Moreover, the group demonstrated that the cytoplasmic ssDNA fragments were newly replicated, implying that they had been released from the stalled fork. Thus, SAMHD1 normally blocks cytoplasmic ssDNA accumulation by preventing the release of ssDNA from stalled replication forks.

Previous work⁵, confirmed by the current study, has shown that SAMHD1 regulates nucleotide levels so that replication forks can progress efficiently across the genome. But Coquel *et al.* found that SAMHD1 also has a second role in fork progression and processing — directly binding to and activating the nuclease enzyme MRE11, which degrades nascent DNA during fork resection⁶ (Fig. 1b). The authors demonstrate that it is this activity of SAMHD1 that prevents newly replicated DNA from accumulating in the cytoplasm. When cells lack SAMHD1, meaning that MRE11 is either absent from stalled forks

or enzymatically inactive, other enzymes aberrantly process the nascent DNA, producing fragments that move to the cytoplasm (Fig. 1c).

Several key steps in this previously undescribed SAMHD1 pathway require further investigation. For example, how does SAMHD1 bind to forks and activate MRE11? What part of the replication fork is processed by SAMHD1 and MRE11 during resection?

Of particular interest is the interplay between SAMHD1, MRE11 and other molecules that control fork resection, such as proteins encoded by the *BRCA1* and *BRCA2* genes, the loss of which promotes certain cancers. *BRCA*-deficient cells exhibit uncontrolled fork resection by MRE11, resulting in over-digested DNA^{7,8}, which could compromise the integrity of the genome and contribute to cancer development. However, this feature of *BRCA*-deficient cells can also make them sensitive to chemotherapy⁸. This treatment induces DNA damage and increases replication-fork stalling — during chemotherapy, *BRCA*-deficient cells might therefore be selectively killed, whereas normal cells are spared. The level of SAMHD1 in *BRCA*-deficient cells might be an indicator of patient responses to chemotherapy, with high levels increasing the activity of MRE11 and

exacerbating replication defects in tumour cells undergoing chemotherapy, and low levels indicating treatment resistance.

SAMHD1 is often mutated in leukaemias and solid tumours^{9,10}. Although altered nucleotide levels might well perturb DNA replication and contribute to tumour development when SAMHD1 is mutated, Coquel and colleagues' research provides an alternative explanation. Fork-resection defects in SAMHD1-deficient cells might lead to increased problems with DNA replication — a phenomenon central to cancer development^{4,6}.

It remains unclear whether cGAS-STING activity, such as that induced by SAMHD1 deficiency, promotes or prevents tumour formation. On one hand, increased levels of cytoplasmic nucleic acids and cGAS-STING activation can signal potentially dangerous replication problems in abnormal cells, and thereby promote their elimination by the immune system⁶. Indeed, cGAS-STING immune signalling is suppressed in some cancers. On the other hand, persistent cGAS-STING signalling can lead to a chronic pro-inflammatory response, which can promote tumour development and spread^{11,12}. Further work is required to better understand the role of these immune responses in cancer.

Coquel *et al.* also advance our understanding of the causes of immune disease by shedding light on the important question of whether the different mutations associated with Aicardi-Goutières syndrome promote disease through a common mechanism. Although cells lacking SAMHD1, TREX1 and RNase H2 all trigger interferon responses through cytoplasmic cGAS-STING signalling, there is some evidence that the activators of this pathway might be distinct. For SAMHD1 and TREX1, there are now links to ssDNA produced during DNA replication^{1,3}, but a role for other cytoplasmic nucleic acids is yet to be ruled out. For RNase H2, cGAS-STING is induced by DNA derived from micronuclei¹³ — small, aberrant nuclei that form when chromosomes fail to segregate properly into sister cells during cell division. Although double-stranded DNA can activate cGAS-STING, ssDNA might also be present in micronuclei and thus contribute to activation of the pathway. It will be interesting to further define exactly which nucleic acids drive this syndrome.

Finally, it remains unclear how nucleic acids are released into the cytoplasm to activate the cGAS-STING pathway. One possibility is that they escape the nucleus after the surrounding nuclear envelope breaks down during cell division¹³. However, Coquel *et al.* found that cytoplasmic ssDNA accumulates rapidly in SAMHD1-deficient cells, even before division, suggesting that other pathways are involved. Understanding how the pathological build-up of nucleic acids in the cytoplasm of cells occurs might help us to identify molecular targets that have the potential to be therapeutically manipulated in immune disease. ■

Madzia P. Crossley and Karlene A. Cimprich are in the Department of Chemical & Systems Biology, Stanford University, Stanford, California 94305, USA.

e-mails: cimprich@stanford.edu;
mpcross@stanford.edu

1. Coquel, F. *et al.* *Nature* **557**, 57–61 (2018).
2. Goldstone, D. C. *et al.* *Nature* **480**, 379–382 (2011).
3. Crow, Y. J. & Manel, A. *Nature Rev. Immunol.* **15**, 429–440 (2015).
4. Zeman, M. K. & Cimprich, K. A. *Nature Cell Biol.* **16**, 2–9 (2014).

5. Franzolin, E. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 14272–14277 (2013).
6. Pasero, P. & Vindigni, A. *Ann. Rev. Genet.* **51**, 477–499 (2017).
7. Schlacher, K. *et al.* *Cell* **145**, 529–542 (2011).
8. Chaudhuri, A. R. *et al.* *Nature* **535**, 382–387 (2016).
9. Clifford, R. *et al.* *Blood* **123**, 1021–1031 (2014).
10. Rentoft, M. *et al.* *Proc. Natl Acad. Sci. USA* **113**, 4723–4728 (2016).
11. Dou, Z. *et al.* *Nature* **550**, 402–406 (2017).
12. Bakhoum, S. F. *et al.* *Nature* **553**, 467–472 (2018).
13. Mackenzie, K. J. *et al.* *Nature* **548**, 461–465 (2017).

This article was published online on 18 April 2018.

ASTRONOMY

Helium discovered in the tail of an exoplanet

As the exoplanet WASP-107b orbits its host star, its atmosphere escapes to form a comet-like tail. Helium atoms detected in the escaping gases give astronomers a powerful tool for investigating exoplanetary atmospheres. [SEE LETTER P.68](#)

DRAKE DEMING

Helium is ubiquitous in the Universe. Large amounts were generated in the Big Bang¹, and nearly every star begins its life by producing helium in its core through the nuclear fusion of hydrogen. The atmospheres of giant exoplanets are expected to have an abundance of helium², because these planets formed from recycled gas and dust from a previous generation of stars. However, searches for helium in such atmospheres have been unsuccessful³. On page 68, Spake *et al.*⁴

report the discovery of helium atoms in the eroding atmosphere of the giant exoplanet WASP-107b. Their work opens a new chapter in the study of exoplanetary atmospheres.

WASP-107b is of comparable size to Jupiter, but has about one-eighth the mass. The exoplanet's low mass relative to its substantial size makes it difficult for the planet to retain its atmosphere — especially in the presence of strong ultraviolet radiation from its host star. Although this star is smaller and cooler than the Sun, it is threaded with magnetic fields produced by the star. Contortions of these

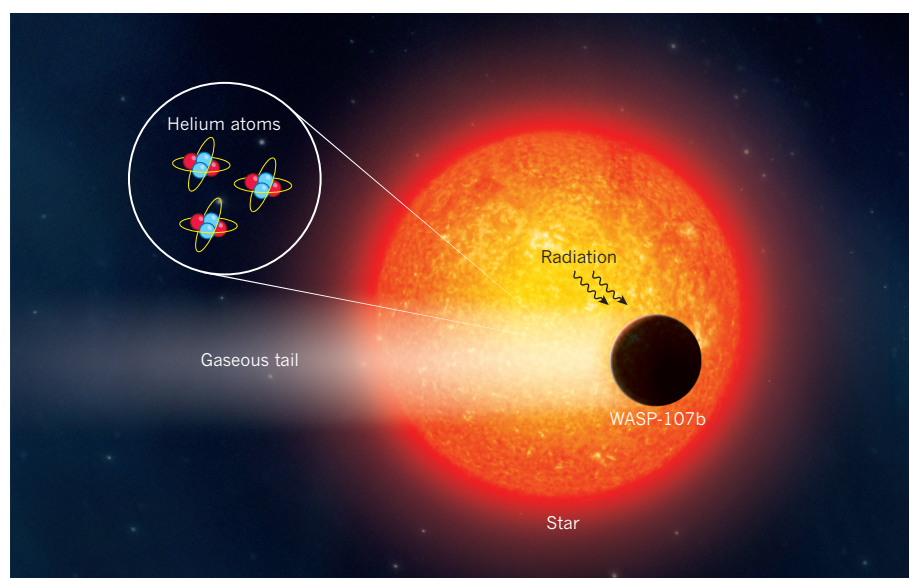


Figure 1 | The escaping atmosphere of WASP-107b. As the giant exoplanet WASP-107b orbits its host star, ultraviolet radiation from the star energizes the planet's atmosphere. Spake *et al.*⁴ show that this causes the atmosphere to escape, and to form a gaseous tail. The authors detected helium atoms in the escaping gases. This is the first time helium has been identified in an exoplanetary atmosphere.

fields emit ultraviolet radiation that energizes the planet's atmosphere.

Spake *et al.* observed WASP-107b using a camera on board the Hubble Space Telescope, and concluded that the planet's atmosphere escapes to form a comet-like tail (Fig. 1). Astronomers have long known that giant planets can lose their atmospheres in this fashion⁵, so this aspect of Spake and colleagues' work is not surprising. But the authors have added a key twist to the story. Until now, only hydrogen (the main component of giant planets) and a few elements with low abundances⁶ have been identified in eroding exoplanetary atmospheres.

Atoms in the gaseous tail of an exoplanet are most easily detected when they absorb stellar light during a transit — a passage of the planet in front of its host star. However, atoms in such a tenuous tail have a tendency to relax to their lowest-energy (ground) state. In this state, most atoms absorb mainly ultraviolet light, and measuring such absorption is difficult for two reasons.

First, Earth's atmosphere is opaque to most ultraviolet light, which means that absorption measurements must be made from space. Currently, only Hubble has the capability for ultraviolet studies of exoplanetary atmospheres, and this telescope could reach the end of its mission lifetime in the next decade. Second, the pattern of how much ultraviolet stellar light is absorbed by transiting planets as a function of time or wavelength tends to be complex. Such complexity makes it difficult to interpret ultraviolet measurements of a transiting planet's atmosphere.

Fortunately, helium atoms have a long-lived (metastable) state, in addition to the ground state. Metastable helium atoms absorb near-infrared stellar light, which has a wavelength only slightly beyond the limits of human vision. Measurements at this wavelength are much easier to interpret than those at ultraviolet wavelengths.

Spake and colleagues observed a transit of WASP-107b, and measured the amount of near-infrared stellar light that was transmitted through the planet's eroding atmosphere as a function of wavelength. The authors identified a narrow absorption feature that they associated with metastable helium atoms (see Fig. 1 of the paper⁴). This signal is more than five times greater than any false signal that could be produced by stellar activity.

Detecting helium in the escaping atmospheres of other exoplanets will be difficult because the absorption signal is intrinsically weak, especially for planets smaller than WASP-107b. However, astronomers will eagerly rise to the challenge. The near-infrared signature of metastable helium is readily transmitted through Earth's atmosphere, which means that eroding exoplanetary atmospheres could be probed using ground-based telescopes. The advent of a new generation of extremely large telescopes at ground-based

observatories⁷ will allow astronomers to study the escaping atmospheres of planets as small as Neptune, which has a radius four times that of Earth.

Theorists have predicted that the atmospheres of Neptune-sized exoplanets could be rich in helium⁸, owing to differences in the rates at which hydrogen and helium are lost to space. Like other giant planets, these bodies are thought to start out with atmospheres of predominantly hydrogen, abundant helium and smaller amounts of elements heavier than helium. As their atmospheres escape, hydrogen is lost fastest, leading to a gradual relative enrichment in the helium content of the atmosphere.

Heavier elements such as carbon and oxygen would be slow to escape, and could in principle be present in exoplanetary atmospheres in concentrated amounts. These heavier elements are key to understanding both how planets form and how they acquire their atmospheres. For planetary astronomers, an escaping atmosphere that is rich in heavy

elements is something of a cosmic treasure, providing ample scientific opportunities to study planetary formation and evolution. Spake and colleagues' detection of helium in WASP-107b will enable astronomers to look for atmospheres that are rich in helium, and perhaps in heavier elements, thereby opening a new subfield of exoplanetary science. ■

Drake Deming is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. e-mail: ddeming@astro.umd.edu

1. Alpher, R. A., Bethe, H. & Gamow, G. *Phys. Rev.* **73**, 803–804 (1948).
2. Seager, S. & Sasselov, D. D. *Astrophys. J.* **537**, 916–921 (2000).
3. Moutou, C., Coustenis, A., Schneider, J., Queloz, D. & Mayor, M. *Astron. Astrophys.* **405**, 341–348 (2003).
4. Spake, J. J. *et al. Nature* **557**, 68–70 (2018).
5. Vidal-Madjar, A. *et al. Nature* **422**, 143–146 (2003).
6. Ben-Jaffel, L. & Ballester, G. E. *Astron. Astrophys.* **553**, A52 (2013).
7. Liske, J., Padovani, P. & Kissler-Patig, M. *Proc. SPIE* **8444**, 84441I (2012).
8. Hu, R., Seager, S. & Yung, Y. L. *Astrophys. J.* **807**, 8 (2015).

PALAEONTOLOGY

Evolutionary insights from an ancient bird

***Ichthyornis dispar* is a key extinct bird species from when birds were shedding characteristics of their dinosaur ancestors and evolving their current features. A reconstructed skull of *I. dispar* now illuminates this transition. SEE LETTER P.96**

KEVIN PADIAN

The distinctive features of birds, from beaks to feathers, provide a stark separation between avians and other animal groups. But how did the features of the bird skull evolve? On page 96, Field *et al.*¹ present a computerized reconstruction of the skull of a pivotal early bird that brings avian evolution into sharper focus.

In the late 1800s, the palaeontologist Othniel C. Marsh and his field crews made many of the first reported discoveries of ancient dinosaurs and mammals from western North America, amassing a fossilized 'bestiary' that dwarfed what was then known from Europe². Marsh's treasures were constantly in the headlines, perhaps never more so than when he published³ his monograph *Odontornithes* in 1880, which reported several previously undescribed fossil birds of the mid-Cretaceous period (around 80 million to 87 million years ago) from the shores of Kansas and nearby states. Familiar yet strange in many ways, these creatures were so archaic that they retained teeth and substantial bony tails, thus providing clues to the reptilian origin of birds. When Charles Darwin

received a copy of the monograph from Marsh, the letter that he wrote back to Marsh said: "Your work on these old birds and on the many fossil animals of N. America has afforded the best support to the theory of evolution, which has appeared within the last 20 years" (see go.nature.com/2hhjxrd).

The specimens Marsh presented in *Odontornithes* were predominantly from two contrasting bird genera: *Hesperornis*, which was flightless and essentially wingless, standing 1.3–1.8 metres tall and comparable to today's loons, and a tern-like bird called *Ichthyornis*, which had an average wingspan of about 60 centimetres (ref. 3). However, neither was closely related to living loons or terns. Both birds had many sharp, curved teeth, which were absent only from the front part of the upper jaw, and their beaks were covered by a horny sheath. Unfortunately, the excavated bones, being small, fragile and of an elaborate architecture, were badly crushed, and proved challenging to prepare. The restoration, mounting and illustration of the specimens were, shall we say, somewhat overenthusiastic. The specimens could be convincingly described only after the mounts had been

disassembled and prepared afresh more than a century later^{1,4}.

Fast forward to the twenty-first century, and in the past 20 years some of the most sensational dinosaur discoveries have been the seemingly endless reports of 'feathered' dinosaurs and newly identified early birds, mainly from Cretaceous deposits in China. These specimens are closer to *Archaeopteryx* (the earliest known bird, from the Late Jurassic of Germany about 145 million years ago) than to *Hesperornis* and *Ichthyornis*⁵. These discoveries have shown that the evolution of feathers, from hair-like down to flight feathers, broadly paralleled the sequence of development of the features of a single feather in living birds⁶. Such insights suggest a plausible sequence for the evolution of wings and flight in birds, whereby newly hatched ancient dinosaurs flapped their incipient wings as a way of boosting their ability to scale steep inclines when evading predators⁷.

But many questions persist about the anatomical changes in early bird evolution, and this is where the work of Field and colleagues comes in. Present-day birds have skulls that are different in many ways from those of all other animals, including the dinosaurs from which they evolved. Bird snouts are lightweight, usually narrow and sometimes quite long. Indeed, the bones of the bird snout are relatively light and fragile compared with those of other animals, and these structures are covered by a strong beak made of the protein keratin, which enables birds to access various foods, such as seeds or carcasses. Inside the beak is a complex of bones that corresponds to the human palate; but unlike ours, the bird bones have mobile connections to each other and to the surrounding skull and jaw bones. This system of mobility is an elaboration of the basic dinosaurian one, and is key to accommodating the diverse feeding habits of birds.

Moreover, 'bird brain' is not the insult you might think. Bird brains are larger relative to their body size than is the case for reptiles, and the relative size of bird brains is comparable to that of placental mammals. As birds evolved from their dinosaur ancestors, the bones that protect the brain enlarged to keep pace with the changes in brain size. The bones of the skull roof and cheek region are also comparatively larger than the equivalent structures in their dinosaur ancestors, whereas the adductor muscles of the bird jaw are reduced. But in what order did these features evolve, and how did they shape avian evolution?

Ichthyornis is closely related to living birds, but retains many features of the earliest birds. No *Ichthyornis* skull material had been uncovered since Marsh's discoveries in the 1870s. But Field *et al.* describe four new three-dimensionally preserved specimens with skull remains, and they image them in 3D, along with some overlooked skull bones from Marsh's original specimens. The authors used a standard technique called high-resolution computed

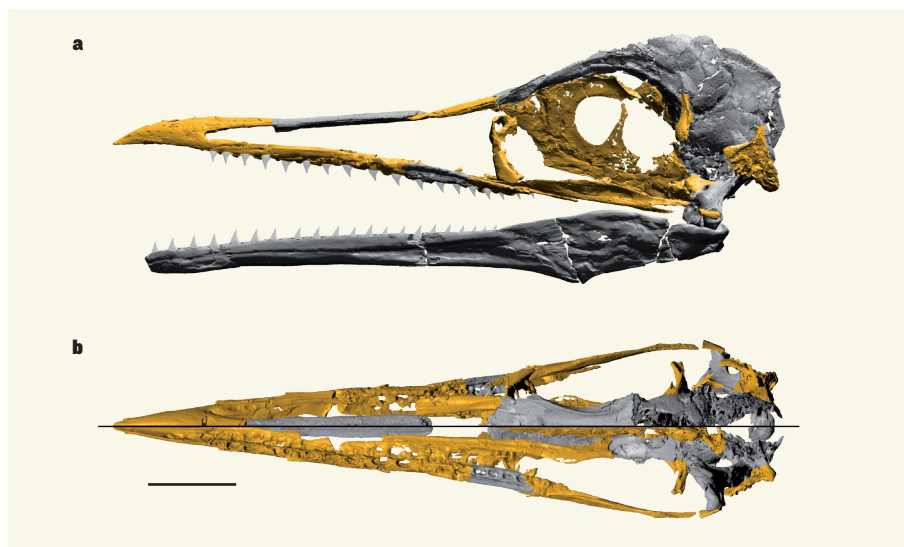


Figure 1 | Skull of the bird *Ichthyornis dispar*. Field *et al.*¹ report the reconstruction of the skull of an extinct species. Their reconstruction fills in some structures missing from previously available fossils, thereby illuminating the transition between the loss of ancient dinosaur features and the evolution of characteristics found in present-day birds. The sections in yellow are newly identified fossil material, whereas the grey structures have been described previously. **a**, A side view of the skull. **b**, A view from above the skull (beak positioned on the left) showing cross-sections in two focal planes. The section above the black line is closer to the top of the skull than the region below the black line. Scale bar, 1 centimetre. (Adapted from Extended Data Fig. 2 of ref. 1.)

tomography, in which a reconstruction of each bone is compiled by taking extremely thin cross-sectional images all the way through the bone, like slicing a salami sausage and reassembling it. This enables the internal anatomy and outer shape to be visualized. The images of all the separate bones are assembled, and a computer program enables the bone images to be manipulated, allowing analysis of how the bones might have moved.

The resulting skull images (Fig. 1) show that the beak of *Ichthyornis* has some features that place it between the earliest birds and living birds: the beak was small, had not yet evolved a bony shelf structure in the palate and was limited to the tip of the jaw. However, the probable mobility of the *Ichthyornis* skull seems to be more like that of living birds. The brain would have been much like those of today's birds, but the cheek region, bounded by bones of the skull roof and the side of the skull, has characteristics that are closer to those of dinosaurs, such as the retention of a large bony chamber for the adductor muscles that close the jaw. Therefore, several key features of the brain and palate evolved before the jaw muscles became reduced and the familiar features of the beak of living birds evolved.

This study raises many questions that remain to be answered. For example, were there functional changes that went along with reducing the jaw muscles from the ancestral dinosaurian condition? Did this change reflect a change in diet? And what ecological habits are correlated with the loss of teeth from the front part of the upper jaw and the evolution of the horny beak that covers it? *Hesperornis* was probably a diver that hunted fishes and

invertebrates in the water column, whereas *Ichthyornis* seems to have been more a surface skimmer or perhaps a shallow plunger like a tern or gull^{2–4}. How did these different predatory approaches favour the same pattern of tooth reduction, which also happened independently in other early bird groups? How did the mobility of the bones of the palate against adjacent skull bones in *Ichthyornis* compare with the ranges of motion in the palates of dinosaurs and living birds, and what might these evolutionary changes suggest about the diet and mode of feeding of *Ichthyornis*?

Whatever the answers to these questions turn out to be, Field and colleagues' beautifully rendered 3D scans and reconstructions of this iconic fossil avian, along with their comparisons of these structures with those of earlier and later birds, provide an important resource to aid our understanding of early bird evolution. ■

Kevin Padian is in the Department of Integrative Biology and the Museum of Paleontology, University of California, Berkeley, Berkeley, California 94720, USA. e-mail: kpadian@berkeley.edu

- Field, D. J. *et al.* *Nature* **557**, 96–100 (2018).
- Plate, R. *The Dinosaur Hunters* (McKay, 1964).
- Marsh, O. C. *Odontornithes: A Monograph on the Extinct Toothed Birds of North America* (US Govt Printing Office, 1880).
- Clarke, J. A. *Bull. Am. Mus. Nat. Hist.* **286**, 1–179 (2004).
- Chiappe, L. M. & Meng, Q. *Birds of Stone: Chinese Avian Fossils from the Age of Dinosaurs* (Johns Hopkins Univ. Press, 2016).
- Prum, R. O. & Brush, A. H. *Sci. Am.* **288**, 84–93 (2003).
- Dial, K. P. *Science* **299**, 402–404 (2003).

A perfect marriage of materials

An absorbent gel has been integrated into the void space of a protein crystal to yield a remarkable self-healing material — it recovers its molecular order after several cycles of expansion and contraction. [SEE LETTER P.86](#)

FRANÇOIS BANEYX

In the best marriages, each partner brings attributes to the union that makes it both stronger and greater than the sum of its parts. On page 86, Zhang *et al.*¹ have achieved such a feat in the field of materials science. They have wed protein crystals to polymeric gels to create a self-healing hybrid material that can undergo multiple cycles of hyperexpansion and contraction without losing its ability to return to its original crystalline state. Such dynamic assemblies hold great promise for applications such as sensors, separators and actuators (devices that convert energy into movement), but have thus far been difficult to produce.

The ordered component of the wedded couple is ferritin, a protein that is ubiquitous in nature. Ferritin stores a mineralized form of iron called ferrihydrite in its hollow core, and releases iron(II) ions (Fe^{2+}) when they are needed by an organism. Each ferritin molecule is made up of 24 subunits that self-assemble to form a nearly spherical, cage-like structure.

Zhang and colleagues used a variant of ferritin, which had been engineered as previously reported², to enable amino-acid side

chains on two ferritin molecules to bind to calcium ions, forming bridges between the molecules. Each ferritin molecule connects in this way to 12 of its neighbours, enabling the protein to grow as cubic crystals that have long-range order extending to tens of micrometres.

The authors wanted to combine the crystal with a flexible component, and chose a polymeric hydrogel — a water-absorbing network of crosslinked polymers — as a suitable candidate. To bring the two components together, the authors first soaked the crystals in a solution of small hydrogel precursors; this allowed the precursors to diffuse throughout the water-filled void space of the ferritin lattice. They then polymerized the precursors by transferring the crystals to a solution of polymerization initiators that was highly saline to prevent undue crystal swelling by water infiltration. A continuous and elastic hydrogel network formed in less than two minutes in the void space, which was infiltrated soon after by the salt solution.

When the resulting composites were placed in water, they expanded isotropically (equally in all directions) to about 200% of their original size within minutes. The crystals did not lose

their faceted morphology, or release ferritin to the bulk medium, for at least 50 minutes. However, continued expansion eventually led to the loss of detectable edges of the crystals.

More strikingly, when Zhang and colleagues transferred partially expanded crystals to a concentrated solution of either sodium chloride or potassium chloride, and then incubated them in a concentrated solution of calcium chloride, the crystals contracted through dehydration, recovering both their original size and their lattice structure (Fig. 1). This expansion–contraction process could be repeated at least eight times without a noticeable change in ferritin ordering. In fact, the authors observed that X-ray structures of ferritin–hydrogel hybrids that had undergone a single expansion–contraction cycle were of higher resolution than could be achieved using conventionally prepared ferritin crystals that lacked the hydrogel. This suggests that the lattices of such hybrids are more precisely ordered than those of conventionally produced ferritin. Polymer infusion might thus be a useful approach to improve the quality of other protein structures, or to access alternative structural states of proteins.

Several additional observations are worth highlighting. First, Zhang *et al.* found that not all hydrogels are equally good at supporting the expansion–contraction process, because extensive electrostatic and hydrogen-bonding interactions are required between polymer side chains and ferritin molecules for order to be recovered. Second, the reversibility of the process is not perfect: calcium ions help the ferritin lattice to ‘snap back’ into place, both by screening undesirable electrostatic interactions and by bridging ferritin molecules, but the structure of about half of these bridges changes after an expansion–contraction cycle. Third, the polymer-infused crystals often cracked when subjected to abrupt expansion or contraction. However, these fractures were swiftly repaired because the polymeric network dynamically interacted with, and rearranged around, the ferritin molecules, even though the hydrogel was not engineered to have intrinsic self-healing properties³.

Finally, the authors showed that neither crystals loaded with ferrihydrite, nor crystals made using fluorescently labelled ferritin, were impaired in their ability to undergo isotropic expansion–contraction when infused with hydrogel. They also demonstrated that the expansion and contraction of such crystals was unaffected when a shell of ferritin was grown on them and the assembly was infused with hydrogel. By contrast, when a ferritin shell was grown on a core of fluorescently labelled ferritin crystals whose molecules had been chemically crosslinked to prevent expansion, and hydrogel was incorporated into the core–shell structure, the shell shattered on treatment of the crystals with water, owing to the generation of a lattice mismatch between the core and shell as the shell expanded.

The internal cavity of ferritin has been used

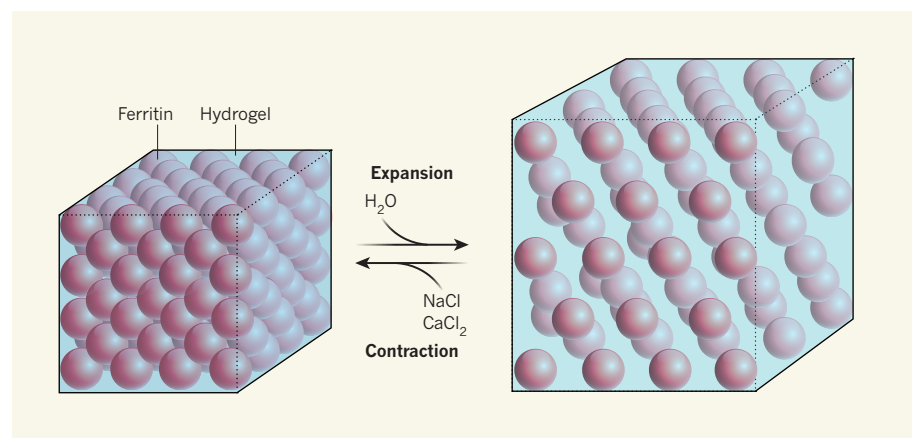


Figure 1 | Expansion and contraction of a protein–hydrogel composite. Zhang *et al.*¹ have prepared a composite material in which a hydrogel (a water-absorbing network of crosslinked polymers) fills the void spaces in the crystal lattice of the almost-spherical ferritin protein; the lattice is held together by calcium ions (not shown) that form bridges between ferritin molecules. When the crystals are soaked in water, they expand equally in all directions as the hydrogel hydrates. Most of the order in the lattice is lost, but some long-range order is retained in the diagonal direction shown (the (111) plane). When the expanded crystals are soaked in a solution of concentrated sodium chloride (which dehydrates the hydrogel) and then in a calcium chloride solution (which re-forms bridges between ferritin molecules), the crystals contract back to their original size, and lattice order is restored.

as a nanoscale reaction vessel in which to produce mineralized nanoparticles that have optical, catalytic or magnetic activity⁴, and packing nanoparticles into ordered assemblies leads to the emergence of collective behaviours that have found applications in opto-electronics, medicine and sensing⁵. Zhang and colleagues' findings might therefore provide a useful method for fine-tuning such collective phenomena by enabling controllable and reversible structural ordering in 3D nanoparticle arrays. Success will hinge on the identity and properties of the particles produced in the cavities of ferritin molecules, on the separation distances and ordering that the protein structure and crystal lattice impose on these nanoparticles, and on whether the kinetics of polymer expansion and contraction can be accelerated or otherwise precisely controlled.

Beyond ferritin, the use of other natural protein cages⁴ — and, more excitingly, of

synthetic protein cages designed from scratch^{6,7} — should provide the versatility needed to control nanoparticle separation distances and lattice systems. Furthermore, the development of creative polymer chemistry will aid efforts to modify the kinetics of expansion and contraction. ■

François Baneyx is in the Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, USA. e-mail: baneyx@uw.edu

1. Zhang, L., Bailey, J. B., Subramanian, R. H. & Tezcan, F. *Nature* **557**, 86–91 (2018).
2. Lawson, D. M. *et al. Nature* **349**, 541–544 (1991).
3. Roy, N., Bruchmann, B. & Lehn, J.-M. *Chem. Soc. Rev.* **44**, 3786–3807 (2015).
4. Rother, M., Nussbaumer, M. G., Renggli, K. & Bruns, N. *Chem. Soc. Rev.* **45**, 6213–6249 (2016).
5. Mann, S. *Nature Mater.* **8**, 781–792 (2009).
6. Bale, J. B. *et al. Science* **353**, 389–394 (2016).
7. Hsia, Y. *et al. Nature* **535**, 136–139 (2016).

NANOTECHNOLOGY

Molecular machines swap rings

A chemical system has been made in which two rings on an axle can switch places by allowing a smaller ring to slip through the cavity of a larger one. The advance opens up potential applications in molecular data storage.

STEVE GOLDUP

Many of the synthetic molecular machines¹ that have been developed in the past 40 years are based on rotaxanes: molecules in which a ring-shaped component encircles a linear axle that is terminated with large 'stoppers' to prevent the ring from slipping off. The threading of the axle through the ring limits the motion of the ring to shuttling back and forth along the axle. Such shuttling has been used in a range of molecular machines that includes switches², ratchets³, pumps⁴ and small-molecule synthesizers⁵. Rotaxanes in which more than one ring encircles the axle have also been made⁶, reminiscent of an abacus, but the rings have been unable to switch places. Writing in *Nature Chemistry*, Zhu *et al.*⁷ now report a system in which the rings can slip past one another, opening the way to new types of molecular machine.

To achieve a ring-through-ring shuttling motion, Zhu and colleagues assembled a rotaxane that contains two differently sized rings (Fig. 1). One has a circumference of 24 atoms, which is about as small as a ring can be in a rotaxane, whereas the other is almost twice as large at 42 atoms. Both rings form hydrogen bonds with nitrogen–hydrogen (N–H) units of

the axle, and this enabled the authors to probe the rings' movement using nuclear magnetic resonance (NMR) spectroscopy.

At room temperature, the authors observed two distinct N–H signals in the NMR spectrum of the rotaxane, because the signal for an N–H unit that is bonded to the small ring appears at a different frequency from that of an N–H unit bonded to the larger ring. This told Zhu and co-workers that the rings exchange places slowly at this temperature, or not at all. However, as the sample of rotaxane was heated, the signals began to broaden and then merged into a single peak. This finding confirmed that the rings change places quickly at elevated temperatures. The only way that this could have occurred is by the smaller ring passing through the larger one.

Zhu *et al.* determined that, at room temperature, the energy barrier that must be overcome for the rings to change places is about 52 kilojoules per mole of rotaxane, which corresponds to a shuttling rate of about 3,600 times per second. For comparison, in an analogous rotaxane that contains only the smaller ring, the ring hops between the N–H groups approximately 80,000 times per second, or roughly 20 times faster. On the basis of this comparison, the authors estimate that the ring-through-ring movement 'costs' about 12 kilojoules per



50 Years Ago

The British General Post Office is busy organizing a "telephone fortnight" in an attempt to silence the public criticism of its services. So far, the promotion has given everybody a chance to tell their favourite telephone stories, most of them unflattering to the GPO. The GPO's timing was inept; it is only two weeks since it announced increases in postal and telephone charges, and it might have been better to let the hubbub settle down before organizing the campaign ... In the next three years, the GPO is intending to spend £1,100 million on investment in telecommunications ... In the longer term, the GPO should be wondering how to increase the number of subscribers ... Britain still has very few telephones — 183 telephones per 1,000 of population.
From *Nature* 4 May 1968

100 Years Ago

Students of animal behaviour will find some interesting facts on the "drumming" of the ruffed grouse ... in *Forest and Stream* for April, illustrated by a series of remarkable photographs, probably the first of the kind which have ever been taken. The author, Mr. F. K. Vreeland, had the good fortune to watch at close range one of these birds while "displaying", and he is convinced that the strange drumming sound then made is produced by the use of the wings alone. This may indeed be the case, but we suspect that later investigations will show that these sounds are at least partly vocal ... The author is apparently so much of an "outdoor naturalist" that he has never read any of the voluminous literature on this theme of courtship displays. But in some respects this adds rather than detracts from the value of his observations, since his records are made without bias.
From *Nature* 2 May 1918

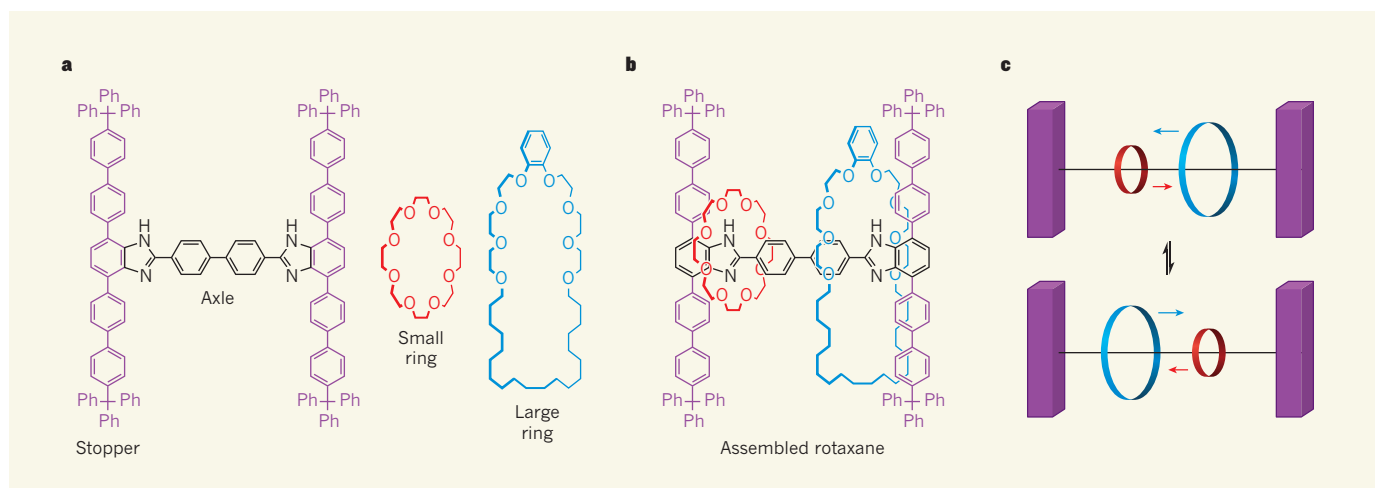


Figure 1 | Ring-through-ring shuttling of a rotaxane. **a, b,** Zhu *et al.*⁷ report a molecule known as a rotaxane that consists of two rings, one much larger than the other, threaded on to an axle. Large groups, known as stoppers, at the ends of the axle prevent the rings from slipping off (Ph, phenyl group). **c,** The authors find that the rings can slip past each other to exchange their positions on the axle. Shuttling takes place by means of the smaller ring passing through the larger one.

mole — a considerable amount, but not as high as might have been expected.

The two-ringed rotaxane is remarkably simple compared with some of the synthetic molecular machines that have been produced so far. It might therefore seem surprising that this is the first time that a ring-through-ring shuttling process has been observed. However, to achieve their breakthrough, Zhu and colleagues had to bring together several key structural features.

First, the dramatic difference in the size of the rings is important for enabling shuttling to occur. Indeed, the authors produced another rotaxane analogue in which the larger ring was 12 atoms smaller and found that no ring-through-ring shuttling occurs, even at elevated temperatures.

Second, when large rings are used as components of rotaxanes, the stoppers on the ends of the axle must be extremely large to prevent the rings from slipping off. This demand can complicate the synthesis of rotaxanes because larger stoppers often cause problems with solubility, and their use typically adds further steps to an already complex synthetic route. Zhu *et al.* overcame these issues by using simple T-shaped stoppers that they had developed previously to make porous materials (known as metal–organic frameworks) that incorporate rotaxanes⁸.

Such structural issues highlight a limitation of the newly identified dynamic process: if rotaxane structures that show ring-through-ring shuttling must be so contrived, will it be possible to use ring-through-ring shuttling to develop molecular machines? It is to be hoped that the answer is ‘yes’, because ring-through-ring shuttling could bring an extra dimension to rotaxane-based switches. A potential application suggested by the authors is molecular-level weaving, in which ring-through-ring shuttling controls the entanglement of molecular threads — essentially, a small ring is used to pull a molecular chain through a larger

ring, akin to threading a macroscopic needle.

Multi-ring rotaxanes could potentially also be used for information storage, in which data are encoded by the order of the rings on the axle. Before now, there was no major advantage to this approach compared with storing data in simpler molecules, because the ring order was fixed at the time of synthesis⁶. Zhu and colleagues’ work opens up the possibility of using external stimuli to order and reorder the rings, and therefore of writing and rewriting any encoded information. ■

Steve Goldup is in the School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK. e-mail: s.goldup@soton.ac.uk

1. Erbas-Cakmak, S., Leigh, D. A., McTernan, C. T. & Nussbaumer, A. L. *Chem. Rev.* **115**, 10081–10206 (2015).
2. Bissell, R. A., Córdova, E., Kaifer, A. E. & Stoddart, J. F. *Nature* **369**, 133–137 (1994).
3. Serrelli, V., Lee, C.-F., Kay, E. R. & Leigh, D. A. *Nature* **445**, 523–527 (2007).
4. Cheng, C. *et al.* *Nature Nanotechnol.* **10**, 547–553 (2015).
5. Lewandowski, B. *et al.* *Science* **339**, 189–193 (2013).
6. Lewis, J. E. M., Winn, J., Cera, L. & Goldup, S. M. *J. Am. Chem. Soc.* **138**, 16329–16336 (2016).
7. Zhu, K., Baggi, G. & Loeb, S. J. *Nature Chem.* <https://doi.org/10.1038/s41557-018-0040-9> (2018).
8. Zhu, K., O’Keefe, C. A., Vukotic, V. N., Schurko, R. W. & Loeb, S. J. *Nature Chem.* **7**, 514–519 (2015).

This article was published online on 30 April 2018.

MEDICAL RESEARCH

Life of a liver awaiting transplantation

People waiting for a liver transplant can die before an organ is found, or, if one is available but of poor quality, there is a risk of transplant failure. A machine that preserves livers might offer a way forward. [SEE ARTICLE P.50](#)

STEFAN SCHNEEBERGER

The standard approach for handling donated livers before transplantation is storing them on ice. On page 50, Nasralla *et al.*¹ report the results of a clinical trial that compared two organ storage methods. More than 200 people who received a liver transplant were randomly allocated either a donor liver that had been stored on ice or one preserved with the aid of a machine that perfuses the organ at body temperature

(37 °C) with oxygenated blood containing nutrients (Fig. 1). The latter method is called normothermic machine perfusion (NMP), and this technique enables organ function to be monitored outside the body before transplantation.

The concept of machine perfusion of an organ awaiting transplantation is not new. Indeed, machine-assisted perfusion was in use before cold storage became the method of choice owing to its simplicity and reproducibility². However, interest in revisiting perfusion

as a transplant approach has been gaining momentum.

The main outcome monitored in the latest trial was the post-transplantation level of the enzyme aspartate transaminase in patients' blood. This measurement is commonly used to assess liver damage and to estimate the risk of transplant failure. The authors found that the use of NMP was associated with less liver damage than that found in livers preserved on ice. Moreover, preservation by NMP reduced the number of organs that were discarded as unsuitable for transplantation compared with livers preserved on ice, and was associated with a better blood-flow profile in the recipient.

The liver's bile ducts can be a point of vulnerability for transplant success, and whether NMP has a positive effect on the viability of these ducts might only be revealed after long-term monitoring. It could therefore be premature to assert that NMP technology has been shown to be a more effective and suitable method for organ storage until additional studies can fully determine the long-term effect of the NMP approach. But it is also fair to say that Nasralla and colleagues' work makes a convincing case for NMP's superiority. Moreover, mimicking the normal conditions for an organ outside the body is a persuasive idea. This clinical trial represents a milestone by directly comparing ice and NMP storage approaches. It could pave the way for the clinical application of NMP and drive a research push in this area.

Organ storage on ice slows the liver's metabolism, and can result in tissue damage by decreasing levels of the energy-storage molecule ATP. The alterations result in accumulation of harmful reactive oxygen species, damage to mitochondrial organelles, and trigger an inflammatory response when blood flow is restored to the transplanted organ³. NMP probably boosts transplant success through mimicking the normal conditions for the organ and enabling ATP replenishment. Such effects would eventually limit the generation of reactive oxygen species and cell damage.

Although Nasralla and colleagues' study is convincing regarding the clinical impact of NMP, it does not clarify the underlying molecular events. Factors including organ architecture and cellular composition will need to be explored in detail to determine the length of preservation times that can be safely used. Other clinical trials of machine-assisted organ perfusion are under way, testing a range of temperatures and conditions⁴. NMP has been performed successfully for 48 hours before transplantation in porcine livers⁵. The median duration of NMP in Nasralla and colleagues' study was around 9 hours.

Pioneering advances in surgical techniques, immunosuppressive drug treatment and patient care have made organ transplantation the standard treatment for chronic organ

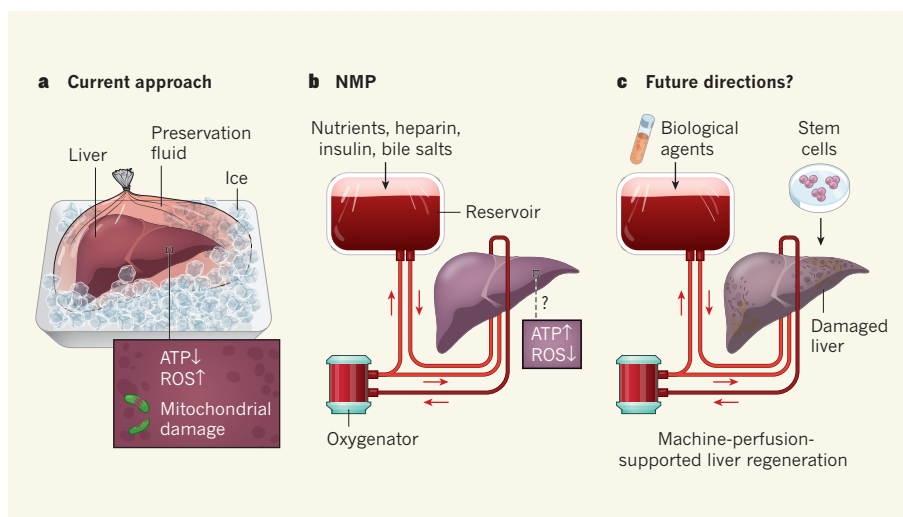


Figure 1 | Supporting the human liver outside the body. Nasralla *et al.*¹ report a clinical trial that compared methods of liver storage before transplantation. **a**, In the current standard approach, livers are stored on ice in preservation fluid. However, this can lead to a decrease in levels of the energy-storage molecule ATP, an increase in harmful reactive oxygen species (ROS) and damage to mitochondrial cellular organelles when blood flow returns after transplantation³. **b**, In an approach termed normothermic machine perfusion (NMP), donated livers are maintained at body temperature (37 °C) by a machine. This device pumps the liver's deoxygenated blood (dark red) through the machine, and then returns oxygenated blood (light red) containing nutrients and essential factors such as bile salts, heparin and insulin to the liver. On the basis of analyses of enzyme levels (not shown), the authors found that NMP results in less liver damage than that incurred when livers are stored on ice. This might be because perfusion results in higher ATP levels and fewer ROS compared with ice storage. **c**, Machine-based organ perfusion could perhaps be adapted to repair liver damage before transplantation. Such an approach might require the addition of biological agents such as growth factors, or the introduction of stem cells.

failure. Yet three fundamental limitations remain: organ shortages; a decline in the quality of donor organs⁶; and restrictions on the time permitted for handling and transporting organs. The use of NMP has the potential to increase the number of organs available for transplantation, and could lessen the need to rapidly transplant an organ after its removal from the donor's body, enabling more time for assessing the liver.

Nevertheless, NMP poses certain challenges. It might be straightforward to set up the machine, prepare the organ for perfusion and do basic technical problem-solving, but substantial training and experience in using this method will be required before it can be routinely used in the clinic. Furthermore, the travel and care plan for an NMP organ is more complex than that for shipment on ice because extra steps are needed in the clinical routine. Standardized protocols and reporting methods need to be established to enable this procedure to advance in a controlled fashion. The sharing of information by medical professionals using NMP could help clinicians gain the collective experience necessary to minimize procedural risks.

Perhaps the greatest advance from this technology is that it might provide proof of principle that organs can survive when a machine helps to mimic the conditions organs encounter in the human body. The human liver itself then becomes like a patient. It is perfused, monitored and fed by a machine. If it performs well, it is transplanted immediately; if not, it

might undergo treatment, modification or repair before transplantation.

The routine clinical use of NMP might boost interest and investment in finding new ways to treat, regenerate and recreate organs that are supported outside the body. The ability to preserve an organ under close-to-normal conditions could be instrumental in advancing not only liver transplantation, but also liver surgery and organ care. Tissue engineering, like transplantation, works towards the goal of providing human organs suitable for transplantation. As NMP offers the potential to modify and monitor an organ, the methods used in both fields might start to merge. This idea is fascinating, because not only would this change many aspects of transplantation, but it might also eventually close the gap between the two technologies. ■

Stefan Schneeberger is in the Department of Visceral, Transplant and Thoracic Surgery, Innsbruck Medical University, 6020 Innsbruck, Austria.
e-mail: stefan.schneeberger@i-med.ac.at

1. Nasralla, D. *et al.* *Nature* **557**, 50–56 (2018).
2. Clark, E. A., Terasaki, P. I., Opelz, G. & Mickey, M. R. *N. Engl. J. Med.* **291**, 1099–1102 (1974).
3. Eltischig, H. K. & Eckle, T. *Nature Med.* **17**, 1391–1401 (2011).
4. Verhoeven, C. J. *et al.* *J. Hepatol.* **61**, 672–684 (2014).
5. Vogel, T. *et al.* *PLoS ONE* **12**, e0188494 (2017).
6. Orman, E. S. *et al.* *Liver Transplant.* **19**, 59–68 (2013).

This article was published online on 18 April 2018.

Genomic variation in 3,010 diverse accessions of Asian cultivated rice

Wensheng Wang^{1,17}, Ramil Mauleon^{2,17}, Zhiqiang Hu^{1,3,17}, Dmytro Chebotarov^{2,17}, Shuaishuai Tai^{4,17}, Zhichao Wu^{1,5,17}, Min Li^{6,7,17}, Tianqing Zheng^{1,17}, Roven Rommel Fuentes^{2,17}, Fan Zhang^{1,17}, Locedie Mansueto^{2,17}, Dario Copetti^{2,8,17}, Millicent Sanciango², Kevin Christian Palis², Jianlong Xu^{1,5,6}, Chen Sun³, Binying Fu^{1,6}, Hongliang Zhang⁹, Yongming Gao^{1,6}, Xiuqin Zhao¹, Fei Shen⁹, Xiao Cui³, Hong Yu¹⁰, Zichao Li⁹, Miaolin Chen³, Jeffrey Detras², Yongli Zhou^{1,6}, Xinyuan Zhang⁵, Yue Zhao³, Dave Kudrna⁸, Chunchao Wang¹, Rui Li³, Ben Jia³, Jinyuan Lu³, Xianchang He³, Zhaotong Dong³, Jiabao Xu⁴, Yanhong Li⁴, Miao Wang⁴, Jianxin Shi³, Jing Li³, Dabing Zhang³, Seunghye Lee⁸, Wushu Hu⁴, Alexander Poliakov¹¹, Inna Dubchak^{11,12}, Victor Jun Ulat², Frances Nikki Borja², John Robert Mendoza¹³, Jauhar Ali², Jing Li³, Qiang Gao⁴, Yongchao Niu⁴, Zhen Yue⁴, Ma. Elizabeth B. Naredo², Jayson Talag⁸, Xueqiang Wang⁹, Jinjie Li⁹, Xiaodong Fang⁴, Ye Yin⁴, Jean-Christophe Glaszmann^{14,15}, Jianwei Zhang⁸, Jiayang Li^{1,10}, Ruairaidh Sackville Hamilton², Rod A. Wing^{2,8*}, Jue Ruan^{5*}, Gengyun Zhang^{4,6*}, Chaochun Wei^{3,16*}, Nickolai Alexandrov^{2*}, Kenneth L. McNally^{2*}, Zhikang Li^{1,6*} & Hei Leung²

Here we analyse genetic variation, population structure and diversity among 3,010 diverse Asian cultivated rice (*Oryza sativa* L.) genomes from the 3,000 Rice Genomes Project. Our results are consistent with the five major groups previously recognized, but also suggest several unreported subpopulations that correlate with geographic location. We identified 29 million single nucleotide polymorphisms, 2.4 million small indels and over 90,000 structural variations that contribute to within- and between-population variation. Using pan-genome analyses, we identified more than 10,000 novel full-length protein-coding genes and a high number of presence-absence variations. The complex patterns of introgression observed in domestication genes are consistent with multiple independent rice domestication events. The public availability of data from the 3,000 Rice Genomes Project provides a resource for rice genomics research and breeding.

Asian cultivated rice is grown worldwide and comprises the staple food for half of the global population. It is envisaged that by the year 2035¹ feeding this growing population will necessitate that an additional 112 million metric tons of rice be produced on a smaller area of land, using less water and under more fluctuating climatic conditions, which will require that future rice cultivars be higher yielding and resilient to multiple abiotic and biotic stresses. The foundation of the continued improvement of rice cultivars is the rich genetic diversity within domesticated populations and wild relatives^{2–4}. For over 2,000 years, two major types of *O. sativa*—*O. sativa* Xian group (here referred to as *Xian*/Indica (XI) and also known as 粳, *Hsien* or Indica) and *O. sativa* Geng Group (here referred to as *Geng*/Japonica (GJ) and also known as 粳, *Keng* or Japonica)—have historically been recognized^{5–7}. Varied degrees of post-reproductive barriers exist between XI and GJ rice accessions⁸; this differentiation between XI and GJ rice types and the presence of different varietal groups are well-documented at isozyme and DNA levels^{6,9}. Two other distinct groups have also been recognized using molecular markers¹⁰; one of these encompasses the Aus, Boro and Rayada ecotypes from Bangladesh and India (which we term the circum-Aus group (cA)) and the other comprises the famous Basmati and Sadri aromatic varieties (which we term the circum-Basmati group (cB)).

Approximately 780,000 rice accessions are available in gene banks worldwide¹¹. To enable the more efficient use of these accessions

in future rice improvement, the Chinese Academy of Agricultural Sciences, BGI-Shenzhen and International Rice Research Institute sequenced over 3,000 rice genomes (3K-RG) as part of the 3,000 Rice Genomes Project¹².

Here we present analyses of genetic variation in the 3K-RG that focus on important aspects of *O. sativa* diversity, single nucleotide polymorphisms (SNPs) and structural variation (deletions, duplications, inversions and translocations). We also construct a species pan-genome consisting of ‘core’ genes that are present in all individuals and ‘distributed’ (variable, accessory or dispensable) genes that are absent in some individuals^{13,14}. The gene presence-absence variations (PAVs) represent another component of species genetic diversity. Our analyses provide new perspectives on rice intra-species diversity and evolutionary history.

Genome mapping, size and SNP variation

Baseline genome sequencing, analyses, and accession information and metadata for the 3,024 rice genomes are summarized in Supplementary Data 1 and Supplementary Notes. Fourteen accessions were excluded from further analyses after quality control. The remaining 3,010 genomes had an average mapping coverage of 92% (74.6–98.7%) (Supplementary Data 2 Table 1), when aligned to the *O. sativa* cv. Nipponbare IRGSP 1.0 reference genome¹⁵ (hereafter referred to as ‘Nipponbare RefSeq’). The estimated size of the genome was

¹Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. ²International Rice Research Institute, Manila, Philippines. ³School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ⁴BGI Genomics, BGI-Shenzhen, Shenzhen, China. ⁵Agricultural Genomics Institute, Chinese Academy of Agricultural Sciences, Shenzhen, China. ⁶Shenzhen Institute for Innovative Breeding, Chinese Academy of Agricultural Sciences, Shenzhen, China. ⁷Anhui Agricultural University, Hefei, China. ⁸Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ, USA. ⁹China Agricultural University, Beijing, China. ¹⁰Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. ¹¹DOE Joint Genome Institute, Walnut Creek, CA, USA. ¹²Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ¹³Advanced Science and Technology Institute, Department of Science and Technology, Quezon City, Philippines. ¹⁴UMR AGAP, CIRAD, Montpellier, France. ¹⁵UMR AGAP, Université de Montpellier, Montpellier, France. ¹⁶Shanghai Center for Bioinformatics Technology, Shanghai, China. ¹⁷These authors contributed equally: Wensheng Wang, Ramil Mauleon, Zhiqiang Hu, Dmytro Chebotarov, Shuaishuai Tai, Zhichao Wu, Min Li, Tianqing Zheng, Roven Rommel Fuentes, Fan Zhang, Locedie Mansueto, Dario Copetti. *e-mail: lizhikang@caas.cn; k.mcnally@irri.org; ccwei@sjtu.edu.cn; ruanjue@caas.cn; nalexandrov@inariag.com; zhanggengyun@genomics.cn; rwing@ag.arizona.edu

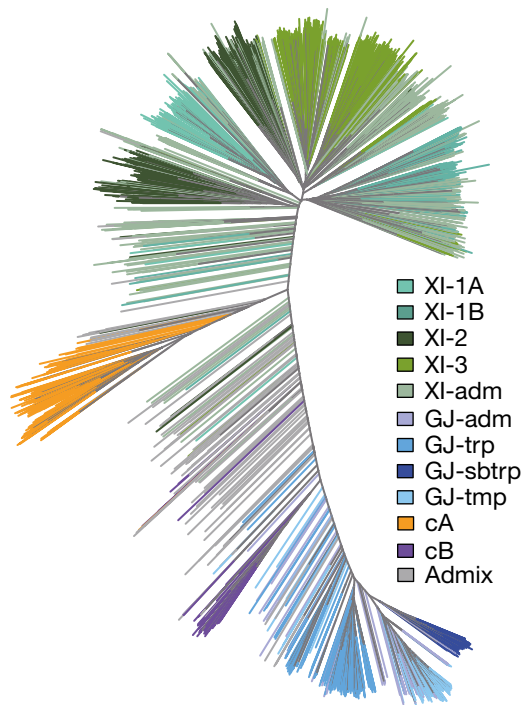


Fig. 1 | Unweighted neighbour-joining tree based on 3,010 samples and computed on a simple matching distance matrix for filtered SNPs. Samples are coloured by their assignment to $k=9$ subpopulations from ADMIXTURE⁴⁶.

375.1 ± 20.9 Mb, with 42.5 ± 0.4% guanine–cytosine content and 35.6 ± 3.7% repetitive sequence content (Supplementary Data 3 Table 1).

We identified over 29 million SNPs—27 million of which are bi-allelic—and found high concordance (>96%) with previous reports (Supplementary Notes)^{16,17}. Filtering reduced this to a ‘base SNP’ set of approximately 17 million SNPs, which captured >99.9% of all SNPs with minor allele frequencies (MAF) > 0.25% (Extended Data Fig. 1). Half (56%) of non-transposable element (NTE) genes and the majority (91%) of transposable element (TE)-related genes have high-effect SNPs (Supplementary Data 2 Tables 2–4). NTE genes contained about 1.44 million moderate-to-high effect, and about 1.5 million low-effect, SNPs, which gave a ratio of 0.95 for moderate-to-high:low SNPs. For small indels, insertions affected 28% of NTE- and 50% of TE-related genes: deletions affected 41% of NTE- and 70% of TE-related genes. A typical genome in a major varietal group contains approximately 2 million (XI and cA), 0.3–0.8 million (GJ; depending on the subpopulation) or about 1.2 million (cB) SNPs (Supplementary Data 2 Table 5). The SNPs of a typical genome were classified as 7.9% moderate-to-high effect and 5.1% low effect.

Population structure and diversity

The 3K-RG accessions were classified into nine subpopulations (Fig. 1 and Extended Data Fig. 2a–d), most of which could be connected to geographic origins (Supplementary Data 1). There were four XI clusters (XI-1A from East Asia, XI-1B of modern varieties of diverse origins, XI-2 from South Asia and XI-3 from Southeast Asia); three GJ clusters (primarily East Asian temperate (named GJ-tmp), Southeast Asian subtropical (named GJ-sbtrp) and Southeast Asian tropical (named GJ-trp)); and single groups for the mostly South Asian cA and cB accessions. Accessions with admixture components <0.65 within XI and GJ were classified as ‘XI-adm’ and ‘GJ-adm’, respectively, and accessions that fell between major groups were classified as admixed (Extended Data Fig. 2a).

Distinct allele frequency profiles for SNPs of MAF > 10% occurred for the nine subpopulations with deviations from the neutral model

reflecting different adaptations and demographic events (Extended Data Fig. 3a). Larger numbers of ‘private’ alleles were found in cA and cB than in other subpopulations (Extended Data Fig. 2e). Comparatively, XI subpopulations have smaller numbers of private alleles, probably owing to ongoing gene flow from natural hybridization and breeding. Doubleton sharing patterns within and between subpopulations showed the same trend (Extended Data Fig. 2f).

Linkage disequilibrium decay rates for combined subpopulations were higher in XI than GJ, with little variation between the two GJ subpopulations, as previously reported^{7,16,18}. However, for the nine subpopulations, linkage disequilibrium decay between XI subpopulations varied more markedly, with XI-2 and XI-3 exhibiting faster linkage disequilibrium decay than XI-1A and XI-1B (Extended Data Fig. 3b). Furthermore, linkage disequilibrium decay correlates strongly with nucleotide diversity (π) among the nine subpopulations ($R^2=0.93$, P value = 2.5×10^{-5}) (Extended Data Fig. 3c).

Nucleotide diversity computation identified many regions of low genetic diversity that contained small numbers of genes under selective constraints (Extended Data Fig. 3d). *Sh4*¹⁹, which controls non-shattering, showed an accordant profile of diversity reduction across all subpopulations (Fig. 2a) that indicates much longer selection, when compared to *qSH1*²⁰. At the semi-dwarf gene *sdl*²¹ locus, a narrow region of reduced diversity occurred in all major groups, which is a similar pattern to that observed for *qSH1*. However, higher diversity in the surrounding 100-kb regions occurred in the cA, cB and XI groups, whereas the GJ groups had extended regions of reduced diversity, which reflects the breeding history associated with the ‘green revolution’²². Different patterns of diversity reduction were observed at other important loci. The *Wx*²³ locus that affects amylose content and stickiness on cooking, the *Badh2*²⁴ locus that affects aroma and their surrounding regions are highly diverse in the XI, cA and cB groups, which indicates complex histories for selection for different types of eating quality; by contrast, both loci and their surrounding regions show low diversity in GJ. The *Rc*²⁵ locus has very low diversity in all variety groups, with variable diversity in the surrounding regions in XI, cA and cB.

We compared SNP variation among TE-related genes, NTE-related genes, 1,021 genes with validated functions curated in the OGRO/QTARO database^{26,27} and a subset of 78 domestication and agronomically relevant genes (Supplementary Data 4). Genetic diversity was reduced significantly (P value < 10^{-12}) near OGRO-curated genes and was often more extreme across the 78-gene subset in each subpopulation (Fig. 2b) when compared with all genomic regions containing genes, which suggests there may have been selection for these genes.

Structural variations

Structural variations (SVs) were called for 3,010 accessions but we focused on 453 accessions with sequencing depths > 20× and mapping depths > 15×, because genome coverage stabilized when sequencing depths exceeded 20× (Extended Data Fig. 4a, b). We identified 93,683 SVs, including 582 SVs larger than 500 kb, with an average of 12,178 SVs per genome. The average sizes of the detected deletions, inversions and duplications are 5.3 ± 0.6 kb, 127.1 ± 19.4 kb and 105.1 ± 22.7 kb, respectively (Fig. 3a, Extended Data Fig. 4c and Supplementary Data 3 Table 2).

SVs showed very strong XI–GJ differentiation. On average, each XI accession differed from Nipponbare RefSeq by 14,754 SVs (8,990 translocations, 5,411 deletions, 188 inversions and 165 duplications), or 3.5× as many as in GJ accessions (Fig. 3a). On average, each cA or cB accession differed from Nipponbare RefSeq by 12,997 SVs and 7,892 SVs, respectively. The total SV sequence that differentiated two GJ accessions was about 22 Mb, whereas it reached 71 Mb between XI and GJ accessions (Fig. 3b). Notably, 1,940 SVs disrupted protein-coding genes within GJ, whereas >6,518 occurred between XI and GJ accessions (Fig. 3c). The SV phylogenetic tree (based on 453 accessions) is similar to the SNP tree, and clearly separates XI, GJ, cA and cB accessions (Fig. 3d). Moreover, the 41,957 major-group-unbalanced SVs that

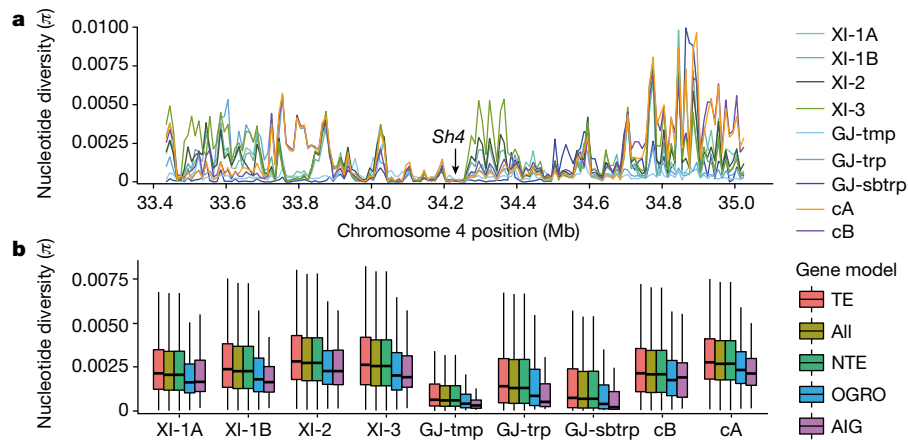


Fig. 2 | Nucleotide diversity. **a**, Differential nucleotide diversity between subpopulations at the *Sh4* locus on chromosome 4 using 10-kb sliding windows. **b**, Box plots of the distribution of π in 100-kb regions surrounding gene models across the genome. Box plots are shown for $k = 9$ subpopulations for all 100-kb windows (All) ($n = 3,728$ in total) and those containing genes annotated as transposable elements (TE) ($n = 3,305$

windows), NTE ($n = 3,709$), from the OGRO/QTARO database (OGRO) ($n = 828$) and the subset of 78 domestication-related genes (AIG) ($n = 61$ windows). Box plots show the median, box edges represent the first and third quartiles, and the whiskers extend to farthest data points within $1.5 \times$ interquartile range outside box edges.

were distributed unevenly among XI, GJ, cA and cB accessions (Fig. 3e) accounted for 44.7% of all SVs and 41.0% of the 582 large SVs.

Pan-genome and population differentiation

The widespread SV and genome size variation (Supplementary Data 3 Tables 1 and 2) encouraged us to investigate the influence of PAVs on protein-coding genes across the 3K-RG. We first used a ‘map-to-pan’ strategy²⁸ to build the species pan-genome (Extended Data Fig. 5a, b), by combining the Nipponbare RefSeq and non-redundant novel de novo assembled sequences; then, PAVs were determined by examining

gene-body and coding sequence (CDS) coverage of mapped reads for each accession.

We identified a total 268-Mb non-redundant novel sequences of length >500 bp with $<90\%$ identity to Nipponbare RefSeq from assemblies of the 3,010 genomes, from which 12,465 novel full-length genes and several thousand novel genes with partial sequences were predicted. Nipponbare RefSeq genes and full-length novel genes could be merged into 23,876 gene families. The *O. sativa* core pan-genome was formed by 12,770 (53.5%) gene families present in all 453 high-coverage genomes, 2,056 (8.6%) without significant gene loss $>1\%$ (P value > 0.05)

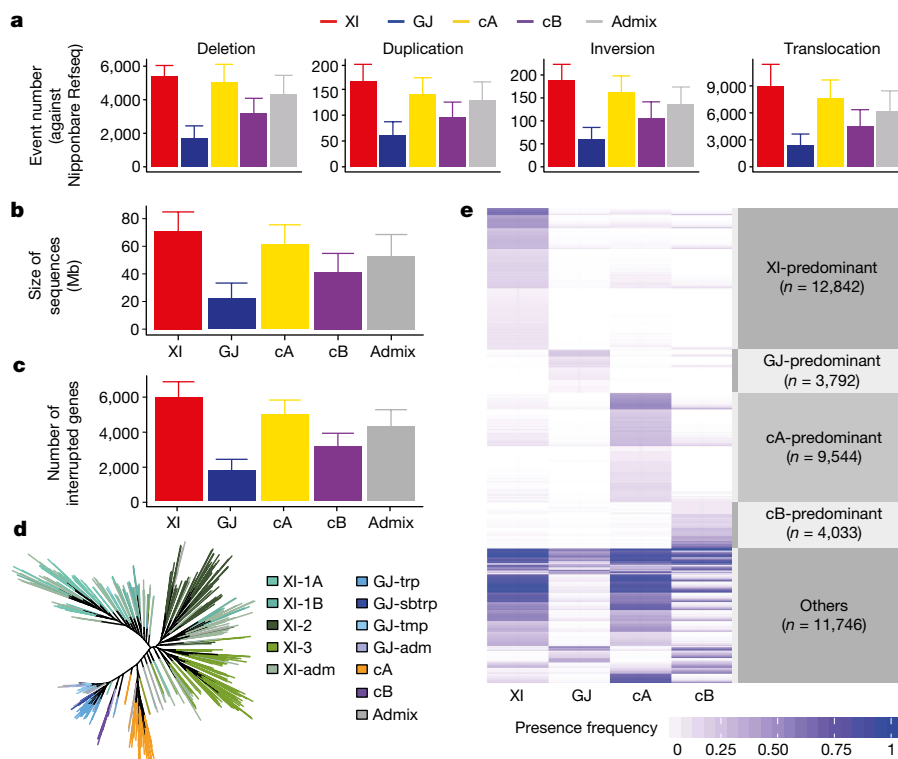


Fig. 3 | Summary of SVs for the 453 high-coverage rice accessions. **a**, Number of deletions, duplications, inversions and translocations. **b**, Genome sizes affected by SVs. **c**, Numbers of genes affected (included or interrupted) by the SVs. **d**, Phylogenetic relationship of 453 rice accessions built from 10,000 randomly selected SVs. **e**, Characterization of the 42,207

major-group-unbalanced SVs unevenly distributed among XI, GJ, cA and cB on the basis of two-sided Fisher’s exact tests. Bar plots in **a–c** are mean \pm s.d. and numbers of accessions in XI, GJ, cA, cB and admix are 303, 92, 33, 10 and 15, respectively.

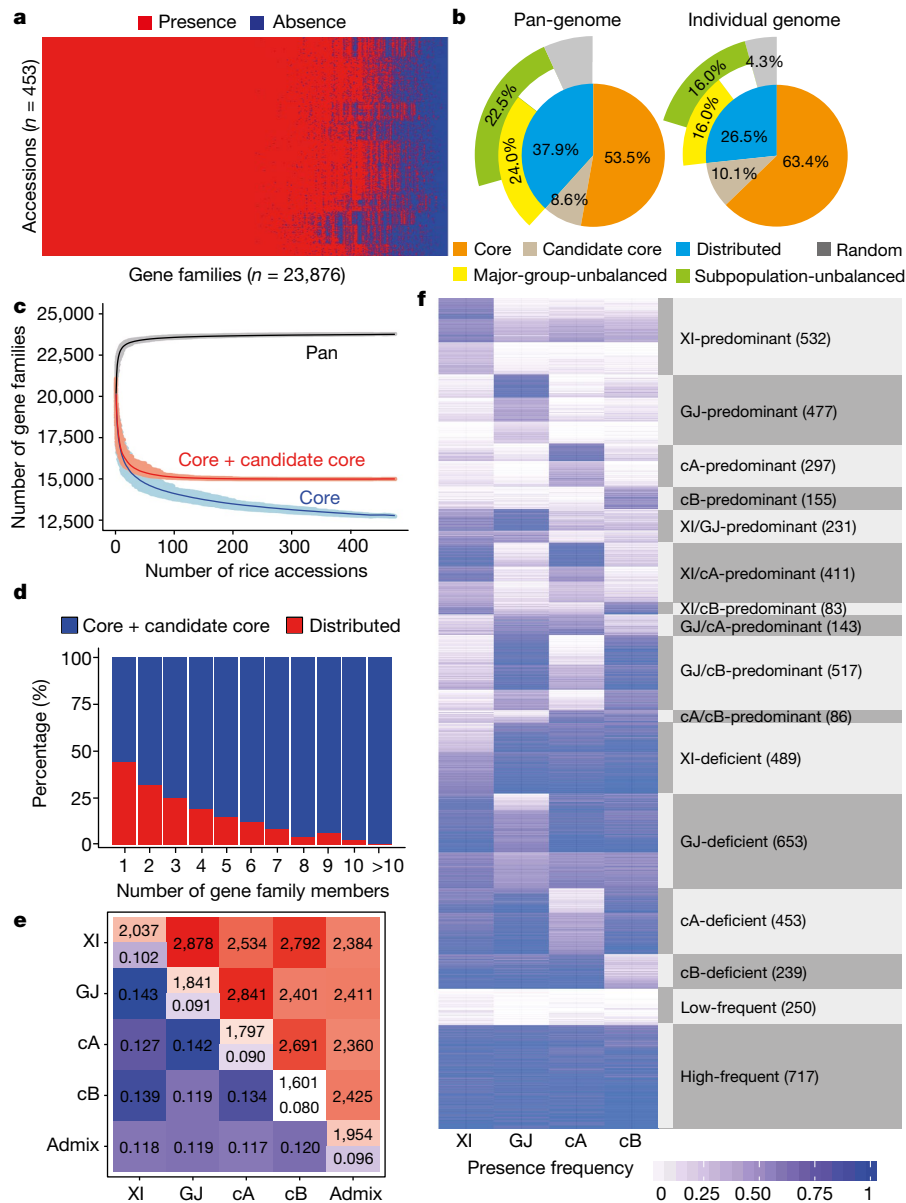


Fig. 4 | Pan-genome of *O. sativa*. **a**, Landscape of gene-family PAVs. Gene families were sorted by their occurrence and rice accessions were clustered with k -means method ($k = 10$). **b**, Compositions of the pan-genome and an individual genome. **c**, Simulation of the pan-genome and core genome based on 500 randomizations of rice genome orders. **d**, Proportions of

the core and distributed gene families binned by gene family sizes. **e**, The average number of gene families that are different between two accessions. **f**, Characterization of 5,733 major-group-unbalanced gene families detected by two-sided Fisher's exact tests.

in all major groups formed candidate core gene families, and the remaining 9,050 (37.9%) comprised distributed gene families (Fig. 4a, b and Supplementary Data 3 Table 3). In silico simulation indicated these 9,050 gene families underestimate the distributed pan-genome (Fig. 4c). Hence, the *O. sativa* pan-genome consists of between 12,770 and approximately 14,826 (53.5% to about 62.1%) core gene families, and at least 9,050 (37.9%) distributed gene families: each accession contains between 63.4% and about 73.5% core gene families and at least 26.5% distributed gene families (Fig. 4b). The core gene families have more members (Fig. 4d) and represent essential gene families. Indeed, 5,476 (36.9%) core or candidate core gene families are enriched in essential functions for growth, development and reproduction (using Gene Ontology, GO), whereas only 862 (9.5%) of the distributed gene families could be annotated with GO terms, showing enrichment in regulation of immune and defence responses and ethylene metabolism (Extended Data Fig. 6a, b).

Pan-genome sequence coverage was evaluated using two new reference genomes²⁹, IR 8 from the XI group and N 22 from the cA group

(Supplementary Data 3 Table 4). We found 98.4% of the IR 8 and 98.6% of the N 22 genome sequences could be mapped to the pan-genome, whereas only 94.3% and 94.0% could be found in Nipponbare RefSeq. By comparing pan-genome data with high-quality XI reference genomes of Zhenshan 97 and Minghui 63³⁰, approximately 25% of the novel genes were shorter owing to gene predictions from fragmented sequences (Extended Data Fig. 5c, d). Novel gene assemblies were validated by mapping raw reads of the 453 high-coverage genomes to the 12,465 novel genes; 11,792 genes (94.6%) had >95% CDS and >85% gene-body coverages were present in at least two rice lines. By comparison, 99.9% of Nipponbare RefSeq annotated genes were detected in the 453 high-coverage genomes (Extended Data Fig. 5e). Approximately 30% of the full-length novel genes were expressed with >1 read per kilobase per million reads in one or more of the 226 publicly available RNA sequencing datasets³¹ (Extended Data Fig. 5f, g). Further, benchmarking universal single-copy orthologues³² evaluation suggested little redundancy in predicted genes (Extended Data Fig. 5h).

Analyses of the PAVs of genes (or gene families) were able to distinguish the major varietal groups, and show that there is considerable variation among and within subpopulations (Extended Data Fig. 7a–d). On average, major group accessions differ by about 4,000 (approximately 10%) genes and about 2,000 (approximately 10%) gene families, whereas XI and GJ accessions differ by more than 6,144 (about 14.9%) genes and 2,878 (14.3%) gene families (Fig. 4e and Extended Data Fig. 7e). The GJ pan-genome has 23,167 gene families comprising 46,115 genes, which makes it 1.9% smaller than XI in terms of gene families and 2.5% smaller in terms of genes. However, all GJ accessions have 240 core gene families (1,594 genes) in common, four times as many as in XI (Extended Data Fig. 7f). In addition, 5,733 major-group-unbalanced gene families were more frequent in some populations but lower in others, including hundreds of XI- and GJ-predominant gene families (Fig. 4f). Moreover, we identified 4,270 XI and 1,384 GJ subpopulation-unbalanced gene families, showing variation between subpopulations within each major group (Extended Data Fig. 7g).

Evolution and domestication of rice

To gain insights into the evolutionary history of the rice pan-genome, gene and gene family ages were estimated by aligning protein sequences to the NR protein database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>) partitioned into 13 taxonomic levels (Extended Data Fig. 8a, b). We observed that: (1) new genes and gene families evolved at alternating rates from phylostratum 1 (PS1) (approximately 3.6 billion years ago) to the emergence of the terminal PS13 clade containing *O. sativa* (about 1.5 million years ago); (2) there was an explosive emergence of new genes accompanying the appearance of *Oryza* at PS12; (3) core genes tended to be more ancient, and most novel genes or gene families were younger and shorter (Extended Data Fig. 8c, d), consistent with recent reports for other species³³; (4) significantly (P value < 0.001) higher SNP variation occurred in distributed genes than in core genes (0.0325 versus 0.0142 SNPs per base) (Extended Data Fig. 8e); and (5) a significantly (P value < 0.001) higher proportion of core genes were under negative selection as compared with those in the Nipponbare RefSeq (Extended Data Fig. 8f).

Regarding *O. sativa* domestication, we constructed haplotype plots for nine important domestication genes—*Rc*²⁵, *Bh4*³⁴, *PROG1*³⁵, *OsC1*³⁶, *Sh4*¹⁹, *Wx*²³, *GS3*³⁷, *qSH1*²⁰ and *qSW5*³⁸ (Fig. 5a–c and Extended Data Fig. 9). Although a large number of XI samples carry an allele found in GJ, many XI accessions carry alleles at each of these loci that are absent in GJ (Fig. 5d). In fact, about 70% of XI accessions do not carry GJ introgressions in at least four genes, and only one XI sample (out of 1,789) had introgressed GJ haplotypes at all nine genes. This observation supports a model of independent domestication of some of the XI pool, rather than the simpler GJ-to-XI introgression hypothesis². Furthermore, the 14-bp deletion in *Rc*²⁵ for domesticated white pericarp was found in several XI lines that carried non-introgressed haplotypes (Extended Data Fig. 9), which suggests independent selection in part of the XI gene pool before introgression of the GJ haplotype became widespread in XI.

Utility of the 3K-RG panel

We demonstrated the use of the 3K-RG genomes and SNPs for trait mapping analyses for the highly heritable traits of grain length, grain width and bacterial blight resistance (Supplementary Notes). Major peaks for grain length with significantly (P value < 10^{-10}) associated markers are on chromosomes 1, 3, 5, 6 and 7, and minor peaks are on chromosomes 4, 9, 10 and 11 (Extended Data Fig. 10a). Major peaks for grain width are found on chromosomes 1 and 5, with minor peaks on chromosomes 3 and 9 (Extended Data Fig. 10b). Genome-wide association study (GWAS) peaks were concordant with known loci, including *GS3*³⁷, *GW5*³⁹, and *qGL7*⁴⁰ for grain length, and *GW5* for grain width. For grain width, the chromosome 9 novel peak coincides with *OsFD1*⁴¹, which codes for a bZIP transcription factor involved in flowering time and developmental plasticity (its pleiotropic regulatory

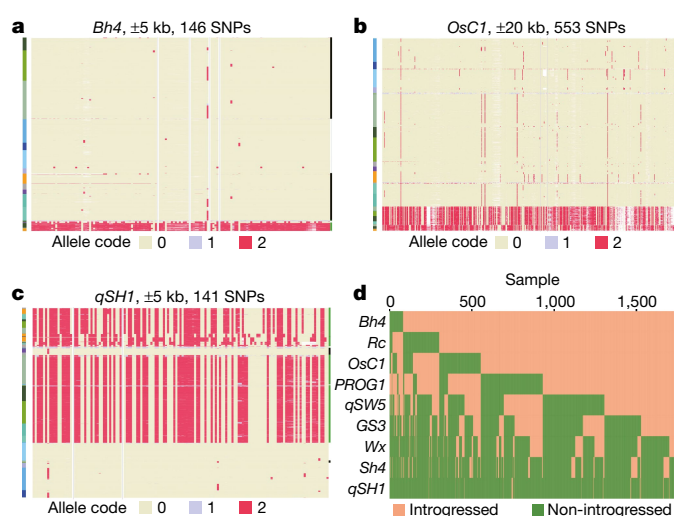


Fig. 5 | Haplotype analyses and introgression. a–c, Haplotypes around the domestication genes *Bh4*(a), *OsC1*(b) and *qSH1*(c). Rows correspond to samples and columns correspond to SNPs. Grey vertical lines mark the gene position. Left colour bar represents the $k=9$ subpopulations. Right colour bar shows introgression status of the XI samples (green, no introgression; black, putative introgression from GJ). d, A heat map showing results of an introgression test of 1,789 XI samples at each of the nine domestication genes. y axis, genes; x axis, XI samples.

function may therefore also affect grain width). Twelve peaks were detected for bacterial blight resistance to strain C5 of *Xanthomonas oryzae*, with the largest clustered around the resistance gene *Xa26*⁴² on chromosome 11 (Extended Data Fig. 10c). Moreover, correlation between gene PAVs and plant height detected the well-known green revolution gene (*sd1*) as the first-ranked candidate. *sd1* is classified as a distributed gene—caused by an approximately 385-bp deletion—and is significantly (P value < 10^{-20}) associated with greatly reduced plant height; it was absent most frequently in XI-1A and XI-1B varieties (Extended Data Fig. 11).

Discussion

We characterized genetic variation in the 3,010 sequenced accessions of *O. sativa* and found a high level of genetic diversity in rice. Although the 3K-RG analysis is expected to identify nearly all polymorphisms with MAF > 1%, our simulations suggest that it includes <40% of rare bi-allelic SNPs (MAF < 1%) in the International Rice Gene bank at the International Rice Research Institute (Extended Data Fig. 1c). We also characterized structural variation, and found that the average number of SVs between pairs of XI genomes (>12,000) was similar to that between two high-quality reference XI genomes³⁰. The vast majority were deletions and translocations distributed across the genome (Extended Data Fig. 4c). Medium-sized SVs (≥ 500 kb) were mostly inversions and duplications, and a large percentage of them (37.9%) occur differentially between XI and GJ. We speculate that large numbers of SVs may contribute to the varying degrees of hybrid sterility and hybrid breakdown between XI and GJ accessions⁴³. We also report pan-genome analyses for *O. sativa*, and the high numbers of PAVs highlight another component of within-species diversity for rice.

Our analysis brings more resolution to the within-species diversity of *O. sativa* (Extended Data Fig. 8e). Larger pan-genomes occur in XI than GJ accessions, but GJ accessions have more core genes than XI (Supplementary Data 3 Table 3), a result that was expected given the greater diversity within XI than GJ. This may relate to differences in eco-geographical distribution: GJ accessions experience harsher high-altitude and/or high-latitude environments, versus the less harsh but more diverse environments experienced by XI rice. Understanding the major group/subpopulation-core, -unbalanced and -predominant gene

functions is expected to shed light on environmental adaptation of rice variety groups over thousands of years.

Although the 3K-RG population structure analyses based on SNPs and SVs were consistent with the five major groups that were previously known, additional subpopulations in the XI and GJ groups were identified and were suggestive of nine subpopulations that are correlated with geographic origin. Large numbers of SNPs, genes and gene families, and SVs were found to be unique to or predominant in single subpopulations. Varying patterns of diversity reduction across different rice subpopulations were observed in and around about 1,000 well-characterized genes. A closer look at patterns of haplotype sharing at domestication genes suggests that not all 'domestication' alleles came to XI from GJ. Taken together, our results—combined with archaeological evidence of XI cultivation for >9,000 years in both India and China^{44,45}—support multiple independent domestications of *O. sativa*.

Our 3K-RG analysis highlights the genetic diversity that exists in rice germplasm repositories, and the usefulness of establishing a digital gene bank in which all accessions can be sequenced and catalogued. For example, we estimate that sequencing the rest of the gene bank of the International Rice Research Institute may enable the identification of >27 million additional SNPs (Extended Data Fig. 1d). The next challenge will be to examine associations of the 3K-RG genetic variation with agriculturally relevant phenotypes measured under multiple field and laboratory environmental conditions; this will guide and accelerate rice breeding by identifying genetic variation that will be useful in breeding efforts and future sustainable agriculture.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0063-9>.

Received: 27 November 2016; Accepted: 28 February 2018;

Published online 25 April 2018.

- Seck, P. A., Diagne, A., Mohanty, S. & Wopereis, M. C. Crops that feed the world 7: rice. *Food Secur.* **4**, 7–24 (2012).
- Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
- Li, L. F., Li, Y. L., Jia, Y., Caicedo, A. L. & Olsen, K. M. Signatures of adaptation in the weedy rice genome. *Nat. Genet.* **49**, 811–814 (2017).
- Wang, H., Vieira, F. G., Crawford, J. E., Chu, C. & Nielsen, R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Res.* **27**, 1029–1038 (2017).
- Ting, Y. Origination of the rice cultivation in China. *J. College of Agric. Sun Yat-Sen University* **7**, 11–24 (1949).
- Glazmann, J.-C. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30 (1987).
- Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S. & McCouch, S. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**, 1631–1638 (2005).
- Chang, T.-T. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. *Euphytica* **25**, 425–441 (1976).
- Han, B. & Xue, Y. Genome-wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant Biol.* **6**, 134–138 (2003).
- Agrama, H. A., Yan, W., Jia, M., Fjellstrom, R. & McClung, A. M. Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. *Nat. Sci.* **2**, 247–291 (2010).
- Allender, C. The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome: Food and Agriculture Organization of the United Nations (2010), 370 pp., ISBN 978-92-5-106534-1. *Exp. Agric.* **47**, (574–574 (2011).
- The 3,000 rice genomes project. The 3,000 rice genomes project. *Gigascience* **3**, 7 (2014).
- Golicz, A. A. et al. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* **7**, 13390 (2016).
- Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154 (2015).
- Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N. Y.)* **6**, 4 (2013).
- Xu, X. et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* **30**, 105–111 (2012).
- Alexandrov, N. et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* **43**, D1023–D1027 (2015).
- McCouch, S. R. et al. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* **7**, 10532 (2016).
- Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
- Konishi, S. et al. An SNP caused loss of seed shattering during rice domestication. *Science* **312**, 1392–1396 (2006).
- Sasaki, A. et al. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**, 701–702 (2002).
- Chandler, R. F. Jr. in *Physiological Aspects of Crop Yield* (ed. Dinanur, R.C.) (Crop Science Society of America, Madison, 1969).
- Wang, Z. Y. et al. The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. *Plant J.* **7**, 613–622 (1995).
- Chen, S. et al. *Badh2*, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *Plant Cell* **20**, 1850–1861 (2008).
- Sweeney, M. T., Thomson, M. J., Pfeil, B. E. & McCouch, S. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **18**, 283–294 (2006).
- Yamamoto, E., Yonemaru, J., Yamamoto, T. & Yano, M. OGRO: The overview of functionally characterized genes in rice online database. *Rice (N. Y.)* **5**, 26 (2012).
- Yonemaru, J.-I. et al. Q-TARO: QTL annotation rice online database. *Rice (N. Y.)* **3**, 194–203 (2010).
- Hu, Z. et al. EUPAN enables pan-genome studies of a large number of eukaryotic genomes. *Bioinformatics* **33**, 2408–2409 (2017).
- Stein, J. C. et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* **50**, 285–296 (2018).
- Zhang, J. et al. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA* **113**, E5163–E5171 (2016).
- The IC4R Project Consortium. Information commons for rice (IC4R). *Nucleic Acids Res.* **44**, D1172–D1180 (2015).
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
- Wilson, B. A., Foy, S. G., Neme, R. & Masel, J. Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth. *Nat. Ecol. Evol.* **1**, 0146 (2017).
- Zhu, B. F. et al. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* **155**, 1301–1311 (2011).
- Tan, L. et al. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **40**, 1360–1364 (2008).
- Saitoh, K., Onishi, K., Mikami, I., Thidar, K. & Sano, Y. Allelic diversification at the C (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* **168**, 997–1007 (2004).
- Fan, C. et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
- Shomura, A. et al. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40**, 1023–1028 (2008).
- Weng, J. et al. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res.* **18**, 1199–1209 (2008).
- Wang, Y. et al. Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat. Genet.* **47**, 944–948 (2015).
- Taoka, K. et al. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* **476**, 332–335 (2011).
- Sun, X. et al. *Xa26*, a gene conferring resistance to *Xanthomonas oryzae* pv. *oryzae* in rice, encodes an LRR receptor kinase-like protein. *Plant J.* **37**, 517–527 (2004).
- Shen, R. et al. Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. *Nat. Commun.* **8**, 1310 (2017).
- Liu, L., Lee, G.-A., Jiang, L. & Zhang, J. Evidence for the early beginning (c. 9000 cal. BP) of rice domestication in China: a response. *Holocene* **17**, 1059–1068 (2007).
- Fuller, D. Q., Allaby, R. G. & Stevens, C. Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. *World Archaeol.* **42**, 13–28 (2010).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

Acknowledgements This work was supported by the 863 Program (2014AA10A601) and National Key R&D Program (2016YFD0100904) from the Chinese Ministry of Science & Technology, Bill & Melinda Gates Foundation Project (OPP1130530), the Agricultural Science and Technology Innovation Program Cooperation and Innovation Mission (CAAS-ZDXT201800), CAAS Innovative Team Award, International S&T Cooperation Program of China (2012DFB32280), Shenzhen Peacock Plan, National Key Technology Support Program (2015BAD01B02), Fundamental Research Funds for Central Non-Profit of CAAS (Y2017CG21) and National Natural Science Foundation of China (31501291, 61472246 and 61272250). IRRI was supported by the CGIAR Research Program CRP 3.3 (Global Rice Science Partnership). We thank the High-Performance Computing Centers at Shanghai Jiao Tong University, Agricultural Genomics Institute of CAAS and State Key Laboratory of Agricultural Genomics, BGI-Shenzhen (2011DQ782025). We thank M. Roa of the Philippine Genome Center Core Facilities for Bioinformatics, Department

of Science and Technology-Advanced Science and Technology Institute of the Philippines, CyVerse, and XSEDE for computing and bioinformatics support; Z. Chong from University of Alabama for help in running novoBreak; the AXA Chair Research Fund and the Bud Antle Endowed Chair for the sequence analysis of IR 8 and N 22 genomes; and Amazon Public Data for free hosting of the 3K-RG analyses results.

Reviewer information Nature thanks M. Bevan, H. Tang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions R.M., D.Ch., S.T., T.Z., R.R.F., F.Z., L.M., M.S., K.C.P., D.Co., J.D., Jiab.X., D.K., S.L., Y.L., M.W., W.H., A.P., I.D., V.J.U., F.N.B., J.R.M., Q.G., Y.N., Z.Y., N.A. and K.L.M. performed the SNP and GWAS analyses. Z.H., W.W., C.S., B.F., X.C., M.C., Y.Zho., Y.Zha., R.L., B.J., Jiny.L., X.H., Z.D., J.S., Jing L., J.T., D.Z., J.A., Jing L., C.We. and Zh.L. performed the pan-genome analyses. Z.W., M.L., W.W., S.T., Jian.X., H.Z., Y.G., X.Zhao, F.S., H.Y., Zi.L., X.Zhan., C.Wa., X.W., Jinj.L., X.F., Y.Y., Zh.L. and J.R. performed de novo genome assembly and the structural variation analyses. D.Co. and R.A.W. conducted all repeat analysis. D.K., J.Z., D.Co., M.E.B.N., J.T., S.L. and R.A.W. sequenced, assembled and annotated the IR 8 and N 22 reference genomes. W.W., R.M., Z.H., D.Ch., S.T., Z.W., M.L., T.Z., R.R.F., F.Z., L.M., D.Co., J.-C.G., Jiay.L., H.L., R.A.W., R.S.H., J.R., G.Z., C.We., N.A., K.L.M. and Zh.L. interpreted data and wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0063-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0063-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Zh.L., K.L.M., C.We., J.R., N.A., G.Z. and R.A.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessments.

Sequencing data of the 3,000 Rice Genome project. The selection and sequencing of rice accessions have previously been described¹². The SNPs/indels and SVs in 3,010 accessions were identified by mapping against the Nipponbare RefSeq, and the pan-genome sequence was created by integrating the Nipponbare RefSeq and non-redundant novel sequences derived from 3,010 rice assemblies. SV comparison and gene PAV analyses focused on 453 rice accessions with sequencing depth $>20\times$ and mapping depth $>15\times$ (Extended Data Figs. 4a, 5b).

Detection of SNPs and indels. Reads were aligned to the Nipponbare RefSeq using BWA-MEM (release 0.7.10)⁴⁷. The mapped reads were then sorted and duplicates were removed by Picard tools (release 1.119) (<http://broadinstitute.github.io/picard/>). The reads around indels were realigned by GATK RealignerTargetCreator and IndelRealigner package (release 3.2-2)⁴⁸. The variants were called for each accession by the GATK UnifiedGenotyper (release 3.2-2)⁴⁸ with 'EMIT-ALL-SITES' option. A joint genotyping step for comprehensive SNP union and filtering step was performed on the 3,010 emit-all-sites VCF files. A variant position is reported if at least one sample supports it with QUAL no less than 30. A total of 29,399,875 SNPs (27,024,796 are bi-allelic) and 2,467,043 indels (small insertions and deletions <40 bp) were identified from the analyses of the genomes of 3,010 accessions. Three subsets of the 3K-RG Nipponbare SNPs were defined using the following filtering criteria: (1) a base SNP set of ~ 17 million SNPs created from the ~ 27 million high-quality bi-allelic SNPs by removing SNPs in which heterozygosity exceeds Hardy–Weinberg expectation for a partially inbred species, with inbreeding coefficient estimated as $1 - H_{\text{obs}}/H_{\text{exp}}$, in which H_{obs} and H_{exp} are the observed and expected heterozygosity, respectively (detailed in Supplementary Notes); (2) a filtered SNP set of ~ 4.8 million SNPs created from the ~ 17 -million-SNP base SNP set by removing SNPs with $>20\%$ missing calls and MAF $< 1\%$; and (3) a core SNP set of SNPs derived from the filtered SNP set using a two-step linkage disequilibrium pruning procedure with PLINK^{49,50}, in which SNPs were removed by linkage disequilibrium pruning with a window size of 10 kb, window step of one SNP and r^2 threshold of 0.8, followed by another round of linkage disequilibrium pruning with a window size of 50 SNPs, window step of one SNP and r^2 threshold of 0.8.

Determining the effects of SNPs. The effects of all bi-allelic SNPs (low, medium and high effects) on the genome were determined based on the pre-built release 7.0 annotation from the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>) using SnpEff⁵¹ release 4.11, with parameters `-v -noLog -canon rice7`. Using sequence ontology terms, a low-effect SNP was classified as 'synonymous_variant', 'splice_region_variant', 'initiator_codon_variant', '5_prime_UTR_premature_start_codon_gain_variant' or 'stop_retained_variant'. A moderate-effect SNP was identified as a 'missense_variant' and a high-effect SNP as a 'start_lost', 'stop_gained', 'stop_lost', 'splice_donor_variant' or 'splice_acceptor_variant'. For indel effects, only indels with lengths that were not multiples of three were counted and SNPs overlapped with protein-coding regions (CDSs of RGAP 7¹⁵ genes) were considered as the most disruptive effects on genes. Results of the SNP and indel effect analysis are given in Supplementary Data 2 Tables 3, 4. We computed the SNP numbers (proportions) of rare SNPs and homozygous singletons for a 'typical genome' of a subpopulation as the median SNP number (proportion) of the SNPs in a given category among those genomes for that subpopulation (Supplementary Data 2 Table 5).

Population structure and SNP diversity. Multi-dimensional scaling analysis was performed using the 'cmdscale' function in R, using the IBS distance matrix of the 3K-RG genomes computed with PLINK^{49,50} on the filtered SNP set. The same distance matrix was used to construct a phylogenetic tree by the unweighted neighbour-joining method, implemented in the R package phangorn⁵². The population structure of the 3K-RG dataset was analysed using ADMIXTURE software⁴⁶ on the core SNP set (version 0.4, <http://snp-seek.irri.org/download.zul>). First, ADMIXTURE was run on 30 random 100,000-SNP subsets of the core SNP set with k (the number of groups) ranging from 5 to 18, and $k=9$ was chosen because it was the minimal value of k to separate all previously known groups (cA, cB, XI, GJ-trp, GJ-tmp and part of GJ-sbtrp). With $k=9$, ADMIXTURE was then run again on the whole core SNP set nine times with varying random seeds; the Q-matrices were aligned using CLUMPP software⁵³ and clustered on the basis of similarity. Then, the matrices belonging to the largest cluster were averaged to produce the final matrix of admixture proportions. Finally, the group membership for each sample was defined by applying the threshold of ≥ 0.65 to this matrix. Samples with admixture components <0.65 were classified as follows. If the sum of components for subpopulations within the major groups XI and GJ was ≥ 0.65 , the samples were classified as XI-adm or GJ-adm, respectively, and the remaining samples were deemed 'fully' admixed (admix). Branches of the phylogenetic tree were coloured according to the $k=9$ admixture classification (Fig. 1).

We computed linkage disequilibrium decay in each subpopulation as follows. The value of r^2 was computed for each pair of SNPs of frequency $\geq 10\%$ in the respective subpopulations that are separated by at most 300 kb using PLINK. The distances were binned into 1-kb bins (separately for each chromosome) and the median value of r^2 in each bin was taken. The medians for each chromosome were then averaged to produce a final r^2 estimate for the bin. We computed nucleotide diversity (π) for non-overlapping 10-kb and 100-kb windows along the Nipponbare RefSeq by adopting an approach similar to VariScan⁵⁴ for genome-wide DNA polymorphism analyses and implemented as a custom R script.

Detection of genomic SVs and population differentiation. Genomic-SV detection for each of the 3,010 rice accessions was performed using a customized version of novoBreak⁵⁵ (<https://sourceforge.net/projects/novobreak/?source=navbar>) against the Nipponbare RefSeq. SVs inferred by no less than 3 reads were further filtered with the following conditions: (1) more than four supporting split reads or (2) no fewer than three discordant read pairs. We detected deletions, inversions and duplications with sizes between 100 bp and 1 Mb, and translocations. Here, translocations were SVs with 'inter-chromosomal breakpoints'. All SVs that passed the filter criteria in the 3K-RG accessions were pooled together. Two adjacent SVs were identified as the same SV if their start and end positions varied no more than 1 kb, and the overlapping region was more than 50% of the total size. The presence-absence matrix of SVs in each accession was built based on this pooled SV dataset. To obtain reliable SV comparison analysis results, we focused only on the 453 high-depth accessions (Extended Data Fig. 4a). Major-group-unbalanced SVs were determined by two-sided Fisher's exact test followed by Benjamini–Hochberg adjustment (false discovery rate (FDR) < 0.05), similar to the detection of major-group-unbalanced genes.

De novo assembly. A variation of SOAPdenovo2⁵⁶ (version r240) with customized k -mers was used to assemble the rice genomes. A k -mer value was initially set for each accession according to a linear model $K=2*\text{int}(0.38*(\text{sequencing depth}+10)+1)$, which was trained from 50 randomly selected rice accessions. The best k -mer value was decided by checking the N50 of the SOAPdenovo results. The command line for SOAPdenovo was 'SOAPdenovo-63mer (or SOAPdenovo-127mer) all -s configure_file (average insertion length set as 460 in the configure file) -K k -mer -R -F' with iteration over different k -mers until N50 of the assembly with that k -mer is higher than those with ' k -mer +2' and ' k -mer -2'. On average, we needed to run SOAPdenovo ~ 3.94 times for each rice accession. The quality of the genome assembly was evaluated for these contigs using QUAST version 2.3⁵⁷.

Sequencing and de novo assembly of IR 8 and N 22 reference genomes. High molecular weight DNA was extracted from young leaves adopting the protocol⁵⁸ with minor modifications. The PacBio library was prepared following the 20-kb protocol (see 'User-Bulletin-Guidelines-for-Preparing-20-kb-SMRTbell-Templates document.pdf', available from https://www.pacb.com/support/documentation/?fwp_documentation_search=PN%20100-286-700-04) and was sequenced on an RSII sequencer with movie collection time of 6 h. The raw data of N 22 and IR 8 were assembled with FALCON⁵⁹ and Canu⁶⁰, respectively. Contigs were polished twice with PacBio raw reads using Quiver (<https://github.com/PacificBiosciences/GenomicConsensus>) and the IR 8 assembly was further polished with $66\times$ WGS $2\times$ 150-bp Illumina data using Pilon⁶¹. Polished contigs were assigned to pseudomolecules using Genome Puzzle Master⁶². Assembly statistics can be found in Supplementary Data 3 Table 4. IR 8 and N 22 were applied to evaluate the completeness and redundancy of the pan-genome.

Pan-genome construction. SOAPdenovo assembly for each accession was assessed by QUAST⁵⁷ with Nipponbare RefSeq as the reference. From QUAST output, unaligned contigs longer than 500 bp were retrieved and merged. CD-HIT version 4.6.1⁶³ was used to remove redundant sequences at a cutoff of 90% identity with the command '`-c 0.9 -T 16 -M 50000`'. For remaining sequences, all-versus-all alignments with BLASTN were carried out to ensure that these sequences had no redundancy. Next, various contaminants including Archaea, bacteria, viruses, fungi and metazoans were removed. The non-redundant sequences were aligned to the NT database (downloaded from NCBI, 26 July 2014) with BLASTN with parameters '`-evalue 1e-5 -best_hit_overhang 0.25 -perc_identity 0.5 -max_target_seqs 10`'. Contigs of which the best alignments (considering E -values and identities) were not from Viridiplantae were considered as contaminants and were filtered out. The remaining contigs formed the non-redundant novel sequences. The rice species pan-genome was then generated by combining the Nipponbare RefSeq and non-redundant novel sequences.

Annotation of the pan-genome. The gene–transcript annotation of the Nipponbare RefSeq was downloaded from the Rice Annotation Project⁶⁴, and if a protein-coding gene contained multiple transcripts only the transcript with the longest open reading frame was selected as the representative for the gene. Protein-coding genes on novel sequences were predicted using MAKER⁶⁵, a gene prediction tool combining ab initio predictions, expression evidence and protein homologies. In detail, repeats were first masked (soft mask for low-complexity

repeats) with RepeatMasker (www.repeatmasker.org) and RepeatRunner⁶⁶. Two ab initio predictors, SNAP⁶⁷ and AUGUSTUS⁶⁸, were called by MAKER⁶⁵ to predict gene models with their default parameters for rice. All rice expressed sequence tags (ESTs) were downloaded from GenBank (15 December 2014) and were aligned to the novel sequences with BLASTN. All rice proteins were downloaded from NCBI (15 December 2014) and were aligned to the novel sequences with BLASTX. To obtain more informative alignments, Exonerate⁶⁹ was used to realign each sequence identified by BLAST around splice sites. EVidenceModeller⁷⁰ was used to combine and refine the ab initio predictions with RNA and protein evidence. Incomplete gene models were removed before the consequent analysis.

Adjustment of predicted genes. We aligned the predicted transcripts against Nipponbare RefSeq to remove potential redundancy. Redundant genes were removed when the genes were clustered into gene families. However, when attempting to identify the number of novel genes, the redundant ones were removed first. We clustered all genes at a global identity of 95%, and removed novel genes that were not representative of the group.

Evaluation of pan-genome redundancy. We ran BUSCO (benchmarking universal single-copy orthologues) v.2.0³² on CX140 (a Nipponbare accession) assembly, Nipponbare RefSeq, CX368 (an N 22 accession) assembly, N 22 high-quality reference genome and the pan-genome sequences. Augustus-3.2.3⁶⁸ and hmmer-3.1b⁷¹ were used for gene prediction in BUSCO. BUSCO was run with genome mode with embryophyta_odb9 as a reference.

Functional analysis. All protein sequences of pan-genome were extracted and aligned to the GO sequence database (<http://geneontology.org/> on 4 April 2015) with BLASTP. Only alignments with E -values $< 1 \times 10^{-5}$ and identity > 0.3 were used. GO terms for each gene were estimated to be the same as those of its best-hit protein. In total, 20,842 (43.3%) genes could be annotated. For a gene family, its GO terms are the non-redundant set of the GO terms of the genes within this gene family. Overall, 6,338 (26.5%) gene families could be annotated. Enrichment of GO terms was carried out using the GStats⁷² package in R with all gene families as the background.

Validation of the non-Nipponbare RefSeq genes. We verified the novel genes by multiple approaches. First, for each gene, we examined the number of accessions that possessed it. We mapped the sequencing reads to the pan-genome sequences. Genes with CDS coverage over 0.95 and gene-body coverage over 0.85 were considered to be present. Second, we verified the novel genes with 226 RNA sequencing experiments from 17 projects³⁰. RNA sequencing reads were first trimmed with Trimmomatic version 0.32⁷³ with parameters 'ILLUMINACLIP:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36' and then aligned to the pan-genome sequences with a split-aligner HISAT2 version 2.0.1-beta⁷⁴ using default parameters. The coverage of each gene was calculated with 'BEDtools coverage' in BEDtools suite version 2.17.0⁷⁵.

Gene family annotation. The genes were clustered to gene families with OrthoMCL version 2.0.9⁷⁶. All genes were extracted and translated into protein sequences and the protein sequences were compared by using all-by-all BLASTP (E -value $= 1 \times 10^{-5}$). OrthoMCL was applied to process the BLASTP output and cluster genes to gene families. Similarity of protein families was set to be 0.5 as suggested by a previous publication⁷⁶.

Determination of gene presence or absence. We proposed a 'map-to-pan' strategy to determine gene presence or absence²⁸. For the 453 accessions with high sequencing depth, although only about 60%–70% of their genomes can be de novo assembled (contig ≥ 500 bp), more than 98% of their genomes can be covered by short read mapping. This enabled the use of coverage of genes to determine their presence or absence. In practice, genes with CDS coverage over 0.95 and gene-body coverage over 0.85 were considered present. If one member of a gene family is present in a given rice accession, the gene family is considered as present.

Determination of core and distributed genes or gene families. A core gene (or gene family) is a gene (or gene family) present in all rice accessions, and we further defined candidate core genes (or gene families) as those with loss rates not significantly larger than 0.01 in all major groups. We first examined whether a gene (or a gene family) is distributed (loss rate > 0.01) in each type of *O. sativa* (XI, GJ, cA and cB). Binomial tests (with a null hypothesis of loss rate < 0.01) were carried out for each gene in each type. A P value below 0.05 meant that this gene (or a gene family) was lost in a significant proportion of rice accessions and is a distributed gene (or gene family) of these subpopulations. If a gene (or a gene family) was not determined to be distributed in all types (and it was not core), it was considered to be a candidate core gene (or gene family) of *O. sativa*. Other genes (or gene families) were considered to be distributed.

Determination of major-group-unbalanced, subpopulation-unbalanced and random genes or gene families. Distributed genes (or gene families) were divided further into major-group-unbalanced, subpopulation-unbalanced and random genes (or gene families). Major-group-unbalanced genes (or gene families) are defined as genes (or gene families) that are unequally distributed among XI, GJ, cA and cB groups. A two-sided Fisher's exact test was used to determine whether

the distribution of each gene (or gene family) is uniform. The P values of all genes were calculated with the 'Fisher.test' function in R and were then adjusted with the Benjamini–Hochberg FDR method. Genes (or gene families) with FDR < 0.05 were considered as major-group-unbalanced.

Subpopulation-unbalanced genes (or gene families) are defined as genes (or gene families) that are unequally distributed among subpopulations; thus, they can be divided into XI-subpopulation-unbalanced genes (or gene families) and GJ-subpopulation-unbalanced genes (or gene families). XI-subpopulation-unbalanced genes (or gene families) are defined as genes (or gene families) that are unequally distributed among XI-1A, XI-1B, XI-2 and XI-3 subpopulations. GJ-subpopulation-unbalanced genes (or gene families) can be defined similarly. The same statistical methods for the major groups were applied to determine the distribution balance for subpopulations. We defined genes (or gene families) that are neither major-group-unbalanced nor subpopulation-unbalanced to be 'random' genes.

Gene and gene-family age. Gene ages were inferred with previously described methods⁷⁷. The NR protein database was downloaded from NCBI (28 March 2015) and all protein sequences were grouped according to 13 taxonomic levels (PS1: Cellular organisms; PS2: Eukaryota; PS3: Viridiplantae; PS4: Streptophyta, Streptophytina; PS5: Embryophyta; PS6: Tracheophyta, Euphyllophyta; PS7: Spermatophyta; PS8: Magnoliophyta, Mesangiospermae; PS9: Liliopsida, Petrosaviidae, Commelinids, Poales; PS10: Poaceae; PS11: BOP clade; PS12: Oryzoideae, Oryzae, *Oryza*; and PS13: *O. sativa*) based on NCBI taxonomy. Thirteen BLASTP databases were built for protein sequences from PS1 to PS13. All genes on pan-genome sequences were first translated into proteins, and were aligned to the 13 databases using BLASTP with E -values $< 1 \times 10^{-5}$ and identity > 0.3 . The age of a gene was considered as the taxonomic level of the oldest aligned protein. Genes that failed to align to all databases were assigned gene ages of PS13 (*O. sativa*). Some PS13 genes were reassigned as PS12 genes if they could be covered by 446 wild rice genomes² with both gene-body coverage > 0.95 and CDS coverage > 0.95 . The age of a gene family was considered as the age of the oldest gene within the gene family.

Introgression test. To test whether an XI sample had a non-introgression haplotype at a locus, we defined a D -value for a sample x as $D(x) = d(x, XI) - d(x, GJ)$, in which $d(x, XI)$ is the mean distance from sample x to a XI sample at the given locus.

With no gene flow from GJ to XI and vice versa, the D -value is negative for XI and positive for GJ. On the other hand, if an XI sample shares a haplotype with a GJ sample, the D -value will be positive and close to the D -values of GJ samples. For an XI sample, we rejected the hypothesis of GJ introgression if its D -value was negative and less than the lower bound of the 99% confidence interval for the D -value of GJ samples, which was computed on the subset of GJ consisting of samples with a positive D -value, to exclude the effect of potential XI-to-GJ introgression.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. Code for studying pan-genome and gene and gene family PAVs are now integrated and published as the EUPAN toolkit²⁸. Tailored novobreak-germline is available at <https://sourceforge.net/projects/novobreak/>; source=navbar. Code for nucleotide diversity and SNP merging is available at <https://github.com/dchebotarov/3k-SNP-paper>. All other code is available from the corresponding authors upon request.

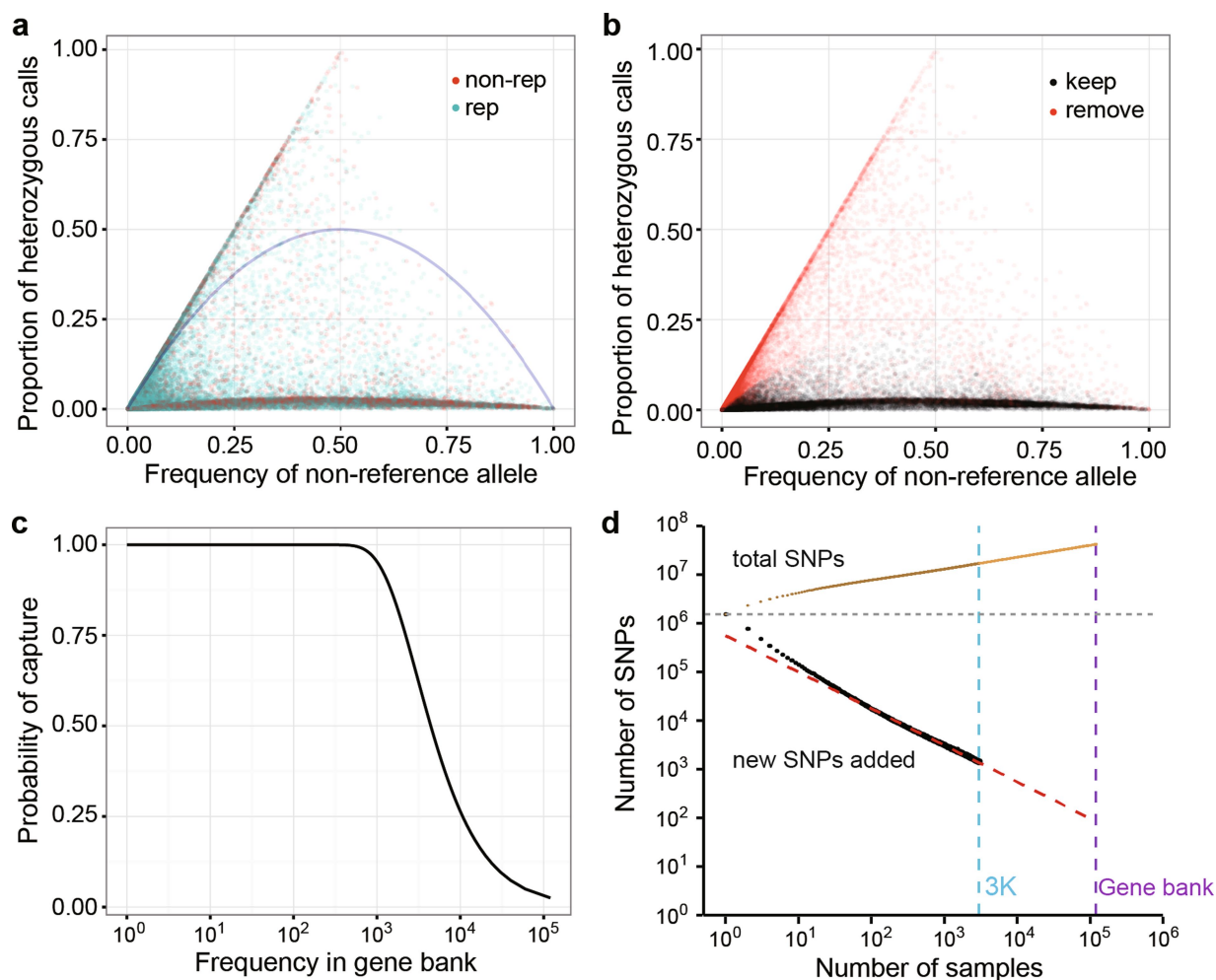
Data availability. The BAM alignment file and variant calls in VCF format for each accession of the 3K-RG against Nipponbare RefSeq are freely downloadable from Amazon Public Data at <https://aws.amazon.com/public-data-sets/3000-rice-genome/> and the Department of Science and Technology Advanced Science and Technology Institute of the Philippines (DOST-ASTI) IRODS site, as described on the 3K-RG project site (<http://iric.irri.org/resources/3000-genomes-project>). The SV and PAV data of 3K-RG are available in the figshare database⁷⁸ (<https://doi.org/10.6084/m9.figshare.c.3876022.v1>).

The following web tools are available for the mining, analysis and visualization of the 3K-RG dataset: SNP-Seek, <http://snp-seek.irri.org>; RMBreeding databases, <http://www.rmbreeding.cn/index.php>; rice cloud of genetic data public projects, <http://www.ricecloud.org/>; IRRI Galaxy, <http://galaxy.irri.org/>; and the 3,000 rice pan-genome browser⁷⁹, <http://cgm.sjtu.edu.cn/3krice/b/>.

The 3K-RG sequencing data used for our analyses can be obtained via project accession PRJEB6180 from NCBI (<https://www.ncbi.nlm.nih.gov/sra/?term=PRJEB6180>), accession ERP005654 from DDBJ (<https://www.ddbj.nig.ac.jp/index-e.html>) and from the GigaScience Database (<https://doi.org/10.5524/200001>).

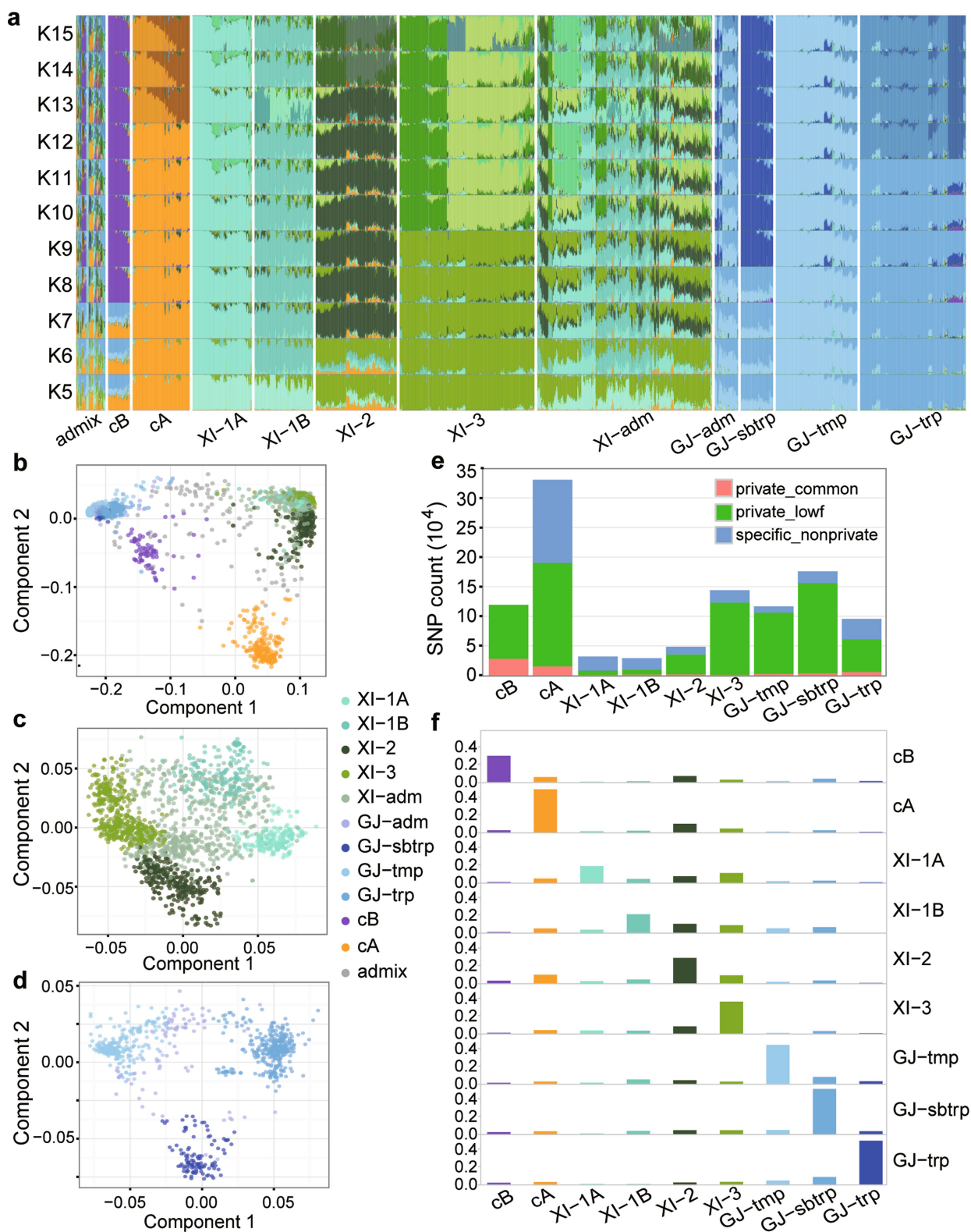
47. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
48. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
49. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

50. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
51. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
52. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
53. Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801–1806 (2007).
54. Hutter, S., Vilella, A. J. & Rozas, J. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* **7**, 409 (2006).
55. Chong, Z. et al. novoBreak: local assembly for breakpoint detection in cancer genomes. *Nat. Methods* **14**, 65–67 (2017).
56. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
57. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
58. Doyle, J. J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
59. Chin, C.-S. et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
60. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
61. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
62. Zhang, J. et al. Genome puzzle master (GPM): an integrated pipeline for building and editing pseudomolecules from fragmented sequences. *Bioinformatics* **32**, 3058–3064 (2016).
63. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
64. Ohyanagi, H. et al. The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.* **34**, D741–D744 (2006).
65. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
66. Smith, C. D. et al. Improved repeat identification and masking in Dipterans. *Gene* **389**, 1–9 (2007).
67. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
68. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
69. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
70. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
71. Finn, R. D. et al. HMMER web server: 2015 update. *Nucleic Acids Res.* **43**, W30–W38 (2015).
72. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
75. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
76. Li, L., Stoeckert, C. J., Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
77. Zhang, Y. E., Vibranovski, M. D., Landback, P., Marais, G. A. & Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* **8**, e1000494 (2010).
78. Hu, Z. et al. Novel sequences, structural variations and gene presence variations of Asian cultivated rice. *Sci. Data* <https://doi.org/10.1038/sdata.2018.79> (2018).
79. Sun, C. et al. RPAN: rice pan-genome browser for ~3,000 rice genomes. *Nucleic Acids Res.* **45**, 597–605 (2017).



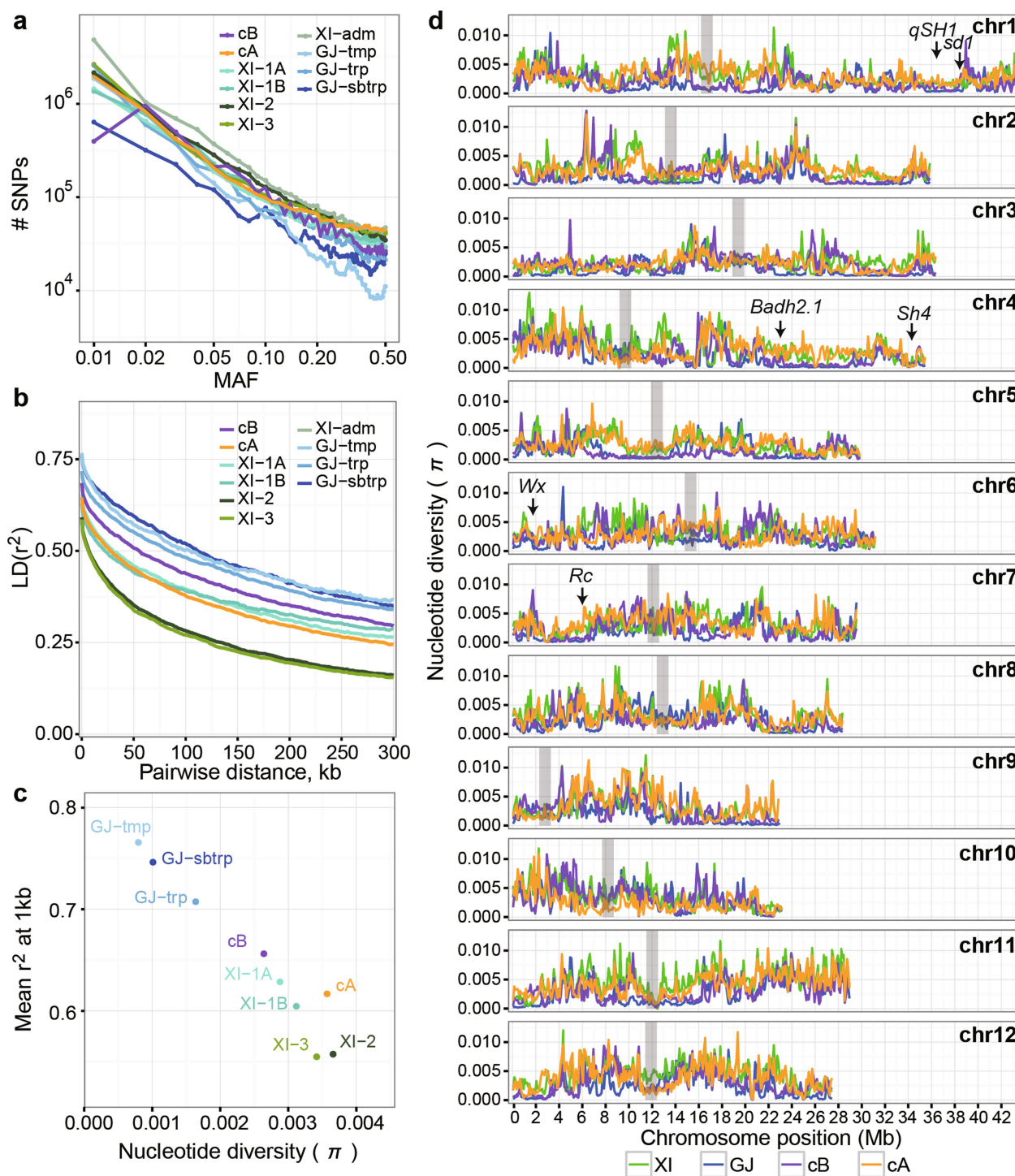
Extended Data Fig. 1 | SNP filtering, discovery rate, and projected discovery upon further sequencing. **a**, Proportion of heterozygous calls versus allele frequency. Each dot represents a SNP from a random sample of 100,000 SNPs. Blue curve shows theoretical Hardy-Weinberg equilibrium. The points have opacity of 5%, such that regions with higher point densities are highlighted. The bulk of SNPs lie on the Hardy-Weinberg equilibrium curve scaled by a factor of about 0.05, which implies a Wright's inbreeding coefficient of $F=0.95$. **b**, The same plot with colour

representing the outcome of filtering. The SNPs that are marked 'keep' (black) form the base SNP set. **c**, The estimated proportion of gene bank SNPs captured by 3K-RG samples, per frequency. The 3,010 samples capture more than 99.99% of gene-bank SNPs of frequency greater than 0.25%. **d**, Projected new SNP discovery rate based on simulations. For a given number of samples (x axis), the graph shows estimated mean number of new SNPs discovered in the last sample.



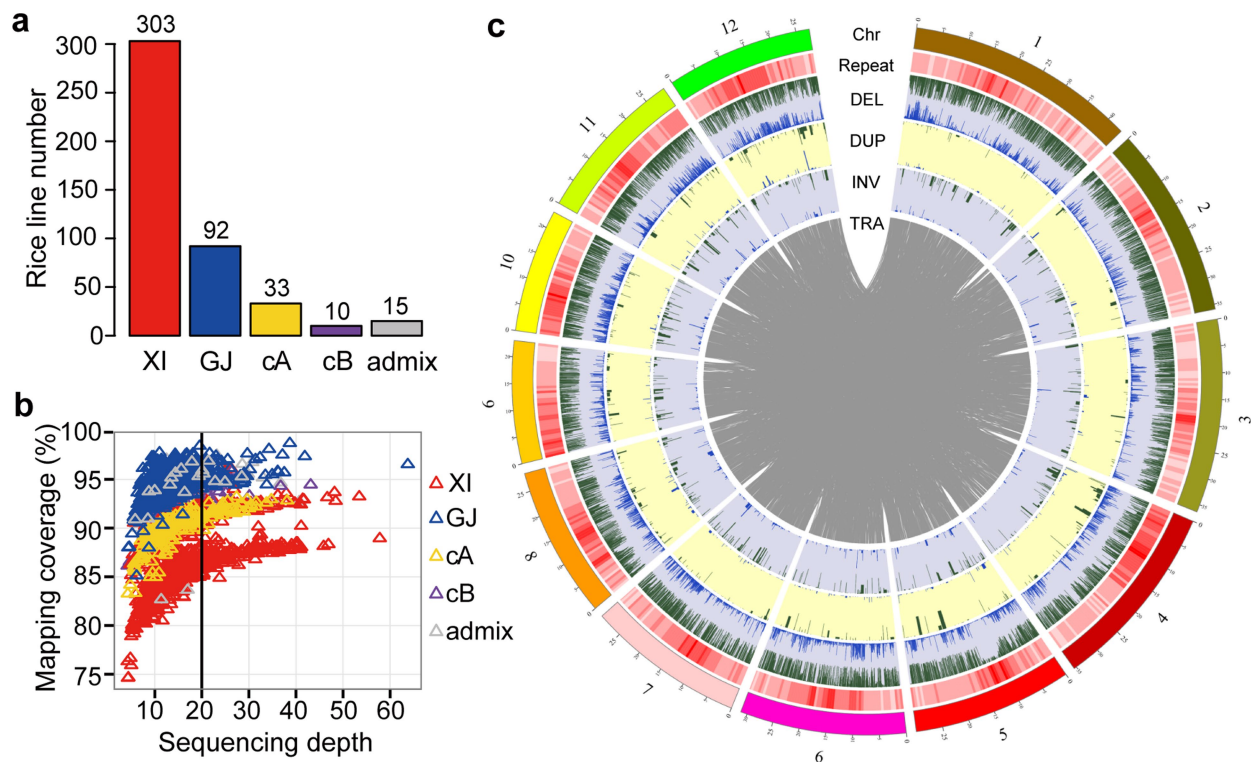
Extended Data Fig. 2 | Population structure and subpopulation differentiation. **a**, ADMIXTURE analyses for $k=5$ to $k=15$. **b–d**, Multidimensional scaling plots for all ($n=3,010$) (**b**), XI ($n=1,786$) (**c**) and GJ ($n=849$) (**d**) accessions. **e**, Private and specific SNPs in each subpopulation. Private alleles are defined as being present in at least

4 accessions in a subpopulation and not found in other subpopulations; population-specific alleles are common in the subpopulation ($\geq 20\%$) but of low frequency ($< 2\%$) in others. **f**, Doubleton sharing—that is, SNPs shared by two accessions—within and between subpopulations, with values normalized by the sample sizes.



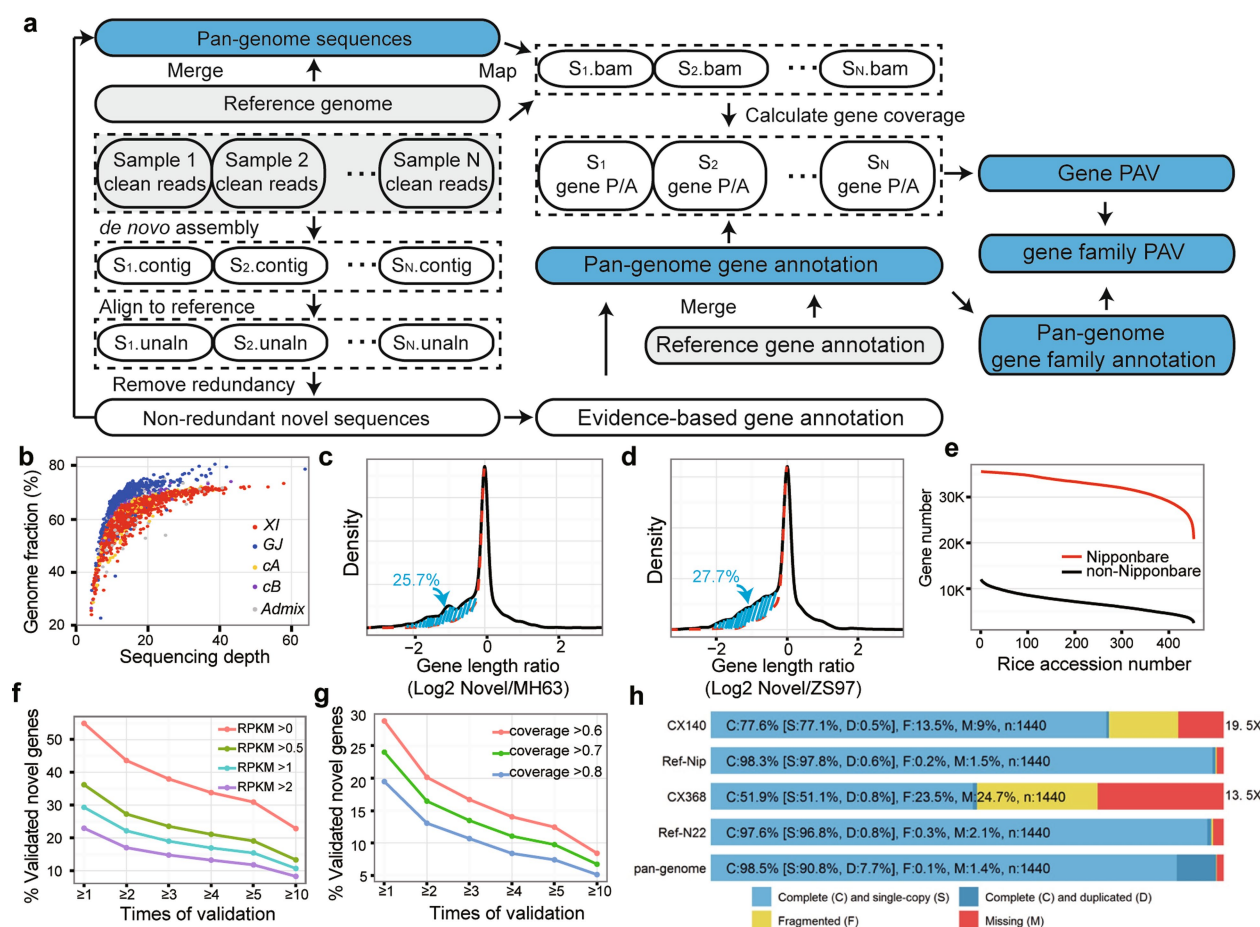
Extended Data Fig. 3 | Genetic diversity within subpopulations. a, MAF histogram. **b,** Genome-wide linkage disequilibrium. **c,** Nucleotide diversity versus linkage disequilibrium. **d,** Diversity scans (π) for all chromosomes

for major groups (XI, GJ, cA and cB) using 100-kb windows in which centromeric regions are highlighted in grey.



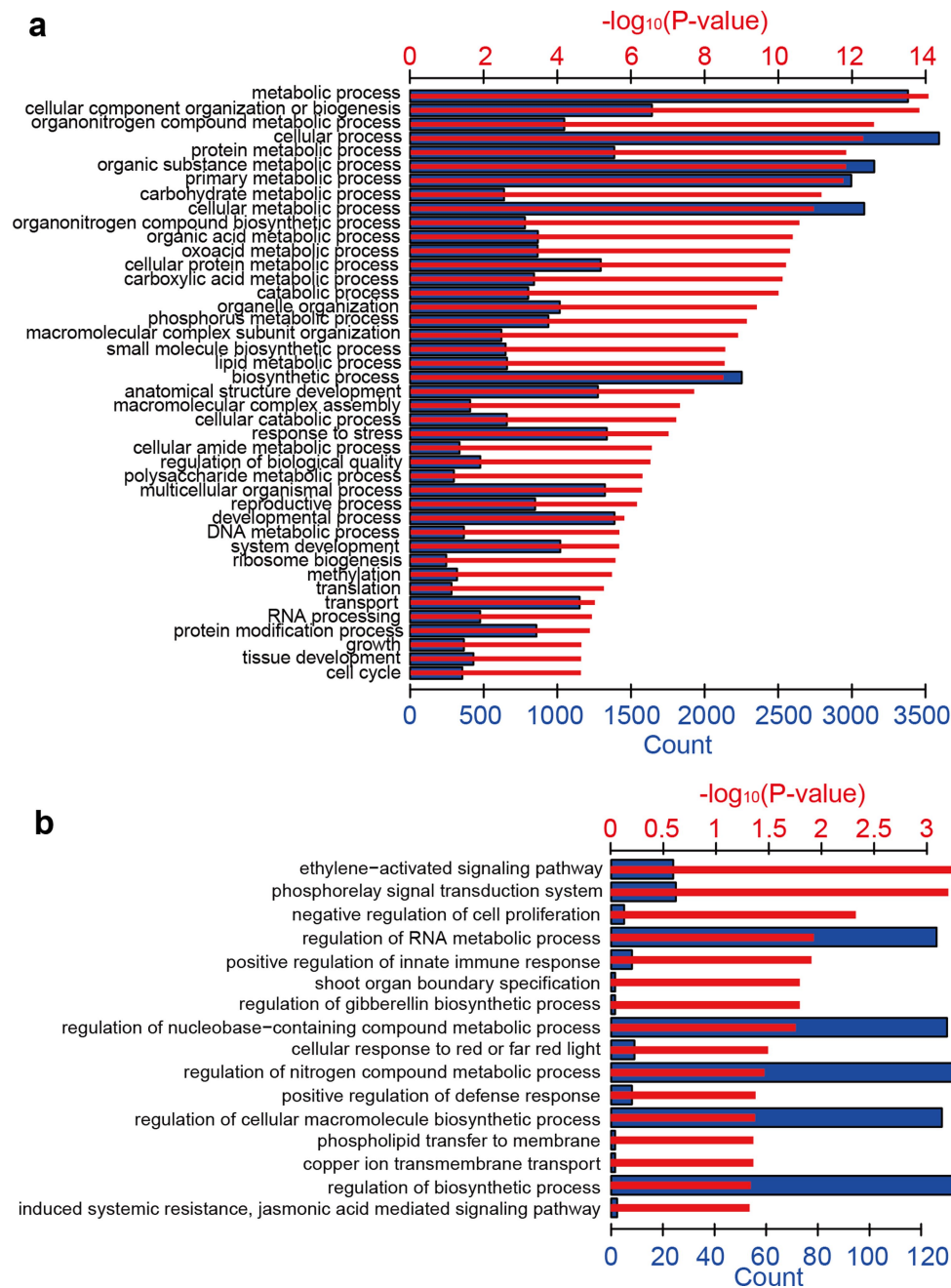
Extended Data Fig. 4 | Selection of high-depth accessions and summary of SVs. a, Number of accessions with sequencing depths $\geq 20\times$ and mapping depth $\geq 15\times$. **b**, Mapping coverage of the 3,010 rice genomes to the Nipponbare RefSeq as a function of sequence depth. **c**, Circular presentation of different types of structural variation detected in 453 high-coverage rice genomes when compared against the Nipponbare RefSeq. Chr, outermost circle represents 12 rice chromosomes with marks in Mb; Repeat, red heat map represents repeat content in 500-kb windows; DEL,

green/blue colour with inner/outer bars represents the average frequencies of deletions detected in XI and GJ; DUP, green/blue colour with inner/outer bars represents the average frequencies of duplications detected in XI and GJ; INV, green/blue colour with inner/outer bars represents the average frequencies of inversions detected in XI and GJ; TRA, grey colour represents translocations across each genome with an average frequency > 0.3 in either XI or GJ.



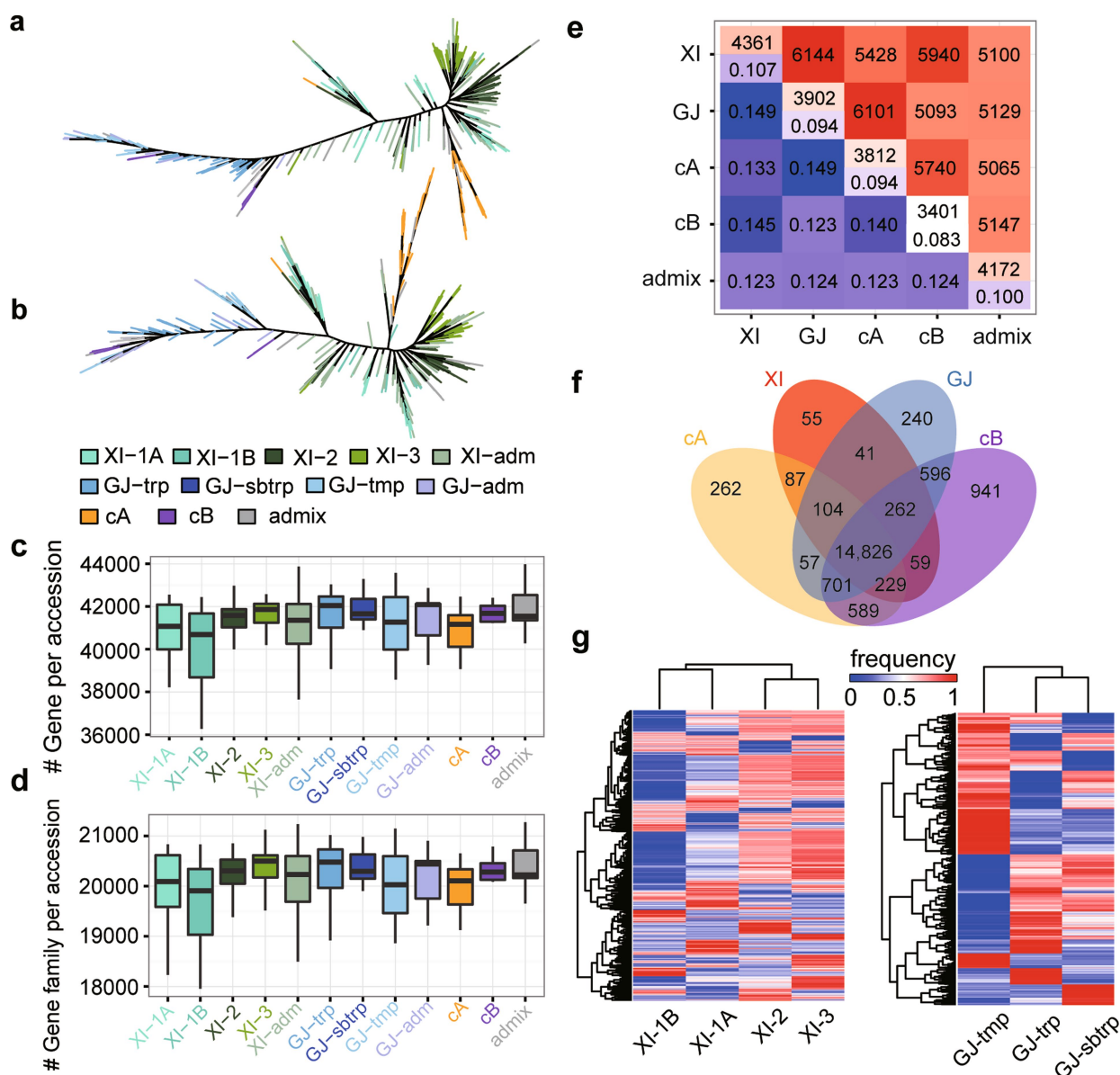
Extended Data Fig. 5 | Map-to-pan strategy for rice pan-genome analyses. **a**, Map-to-pan pipeline for pan-genome analyses: (1) pan-genome sequences were derived by combining Nipponbare RefSeq and de novo assembled non-redundant novel sequences; (2) gene annotations were derived by combining Nipponbare RefSeq annotations and evidence-based gene predictions on novel sequences; (3) reads from each sample were mapped to pan-genome sequences; and (4) gene presence or absence was determined by coverage of mapped reads. Raw data for this pipeline is shown in grey boxes and the main output is shown in blue. P/A: presence/absence. **b**, Proportions of assembled genomes as a function of the sequencing depth (based on the Nipponbare RefSeq). **c**, **d**, Gene-length differences between the novel pan-genome genes and genes derived from the genome of Minghui 63 (MH63) (**c**) or Zhenshan 97 (ZS97) (**d**). Generally, the distribution should be symmetric: a ratio of > 0 means the

novel gene is longer and a ratio < 0 means the novel gene is shorter. The dashed red lines show the symmetric distributions of the >0 part and the blue regions show the gene proportion with shorter lengths. **e–g**, Genomic (**e**) and transcriptomic (**f**, **g**) validation of novel genes. **e**, Validation based on genomic sequencing data, in which numbers of the Nipponbare RefSeq and non-Nipponbare RefSeq genes identified (>95% CDS coverage and >85% gene-body coverage) are shown against the numbers of supporting rice accessions in the 453 rice lines; **f**, **g**, Validation based on the mapping rates of the publicly available RNA sequencing data of rice, including gene expression (**f**) and coverage of the coding sequence (**g**). **h**, BUSCO evaluation for 1,440 highly conserved genes; CX140, the assembly of Illumina sequencing data of Nipponbare accessions; Ref-Nip, Nipponbare RefSeq. CX368, the assembly of Illumina sequencing data of accession N 22; Ref-N22, assembly of N 22 PacBio sequencing data.



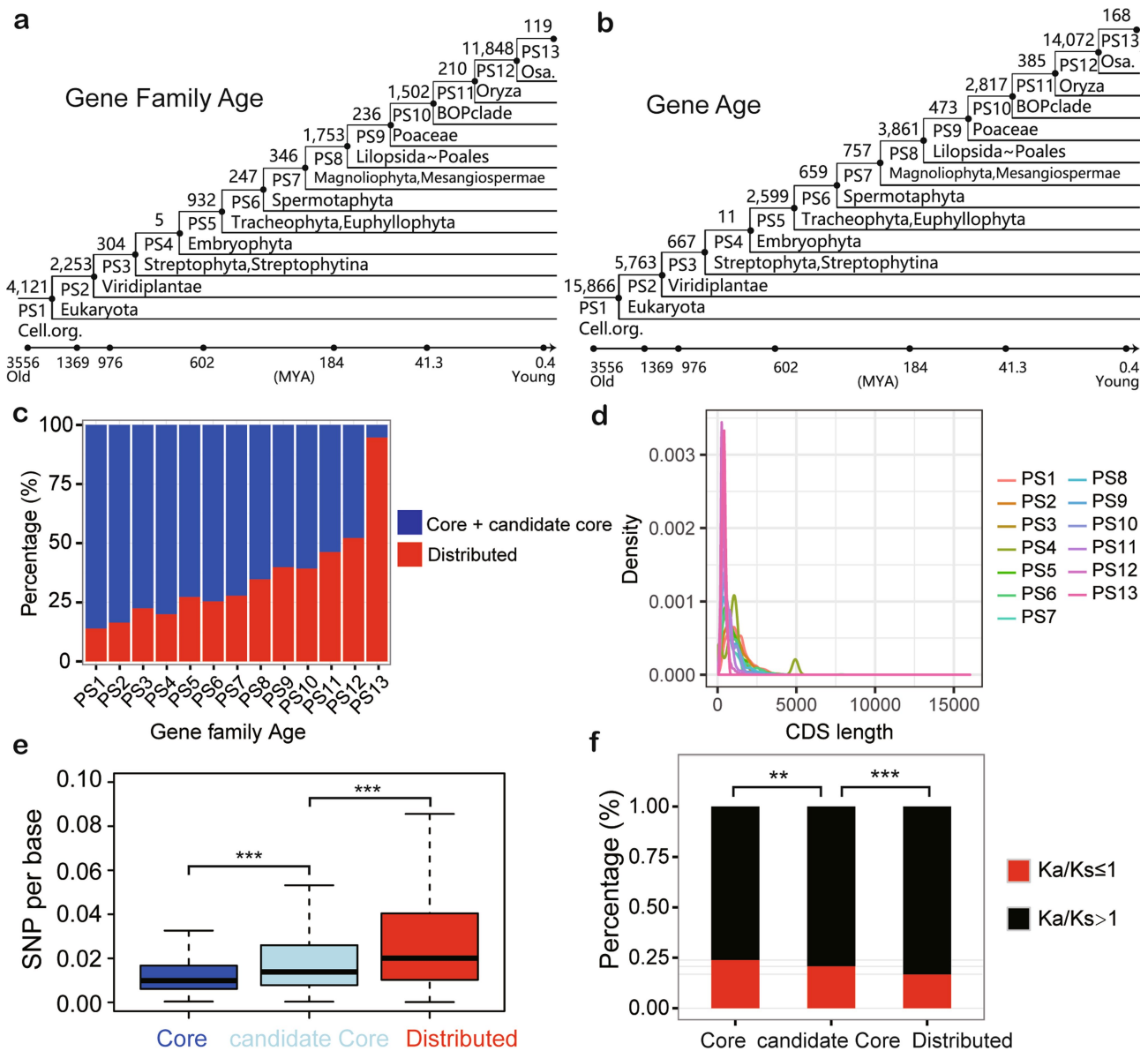
Extended Data Fig. 6 | Representative enriched biological processes of core and distributed gene families. a, b, Representative enriched biological processes of core (a) and distributed gene families (b) are shown, with all terms sorted by their enriched P values (red bars).

One-sided hypergeometric test built in the GOSTATS R package was used to calculate the P value of each GO term. The numbers of gene families involved in each GO term are shown in blue.



Extended Data Fig. 7 | Characterization of gene or gene family presence or absence variations. **a, b**, Phylogenetic trees of the 453 rice accessions constructed on the basis of the presence or absence of the distributed genes (**a**) and gene families (**b**); both of which classified the 453 accessions into two major groups (XI and GJ), with each being further divided into several subpopulations that are tagged with different colours representing their classifications based on the SNP. **c, d**, Gene (**c**) or gene family (**d**) numbers per accession in different subpopulations; gene or gene family numbers were significantly different among XI subpopulations (Kruskal–Wallis tests, P value = 9.8×10^{-8} (gene) or 1.0×10^{-6} (gene family)). Box plots

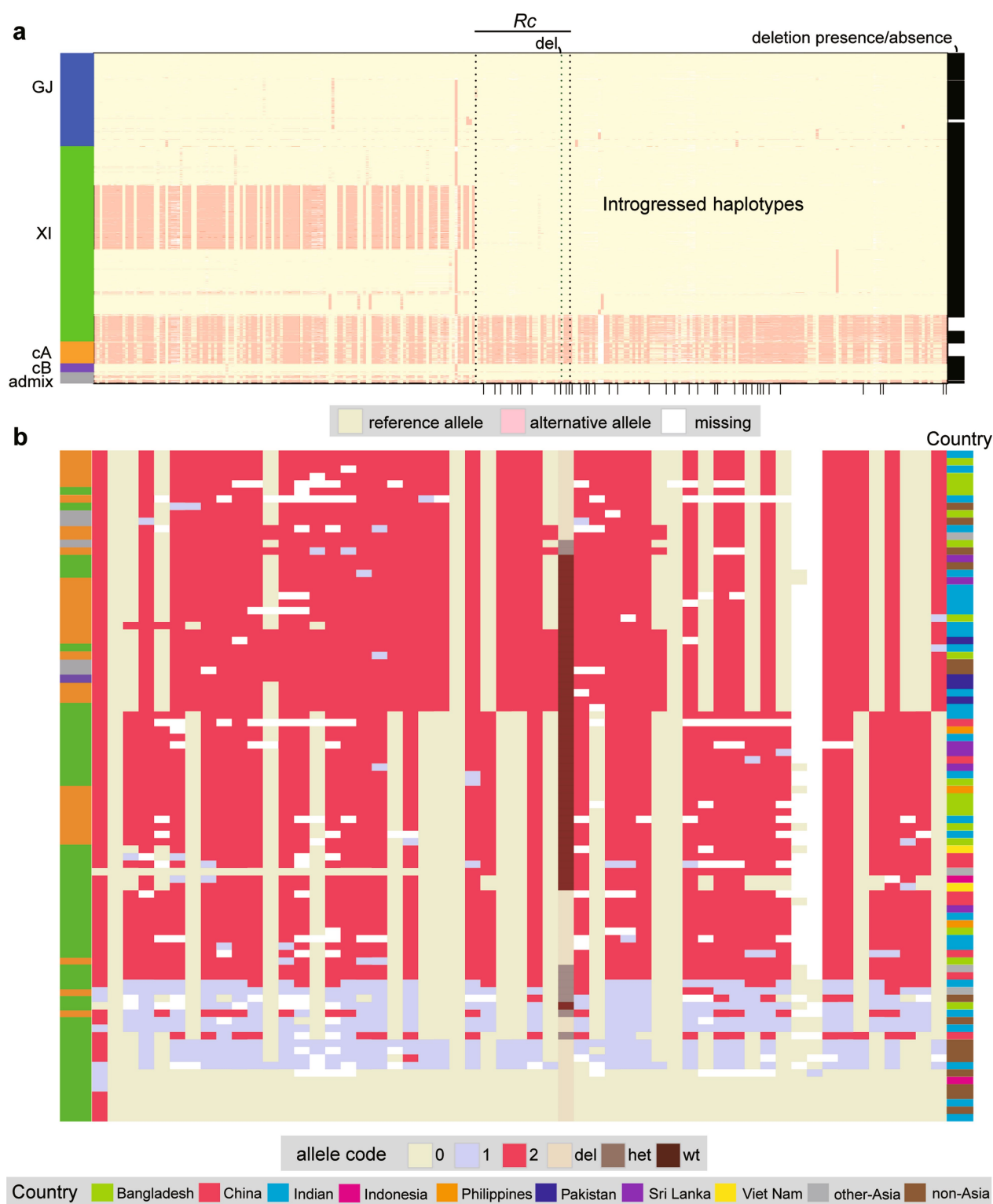
show the median, box edges represent the first and third quartiles and the whiskers extend to $1.5 \times$ interquartile range. **e**, The average number of genes that are different between two accessions in which all combinations of the 453 accessions were considered, and the proportions were calculated as the number of such differentiating genes adjusted by the gene numbers held in common by the two genome types. **f**, Venn diagram of the numbers of the core + candidate core gene families among the major groups of *O. sativa*. **g**, Cluster analysis of 4,270 XI-subpopulation-unbalanced gene families and 1,384 GJ-subpopulation-unbalanced gene families.



Extended Data Fig. 8 | Evolution of the pan-genome of *O. sativa*.

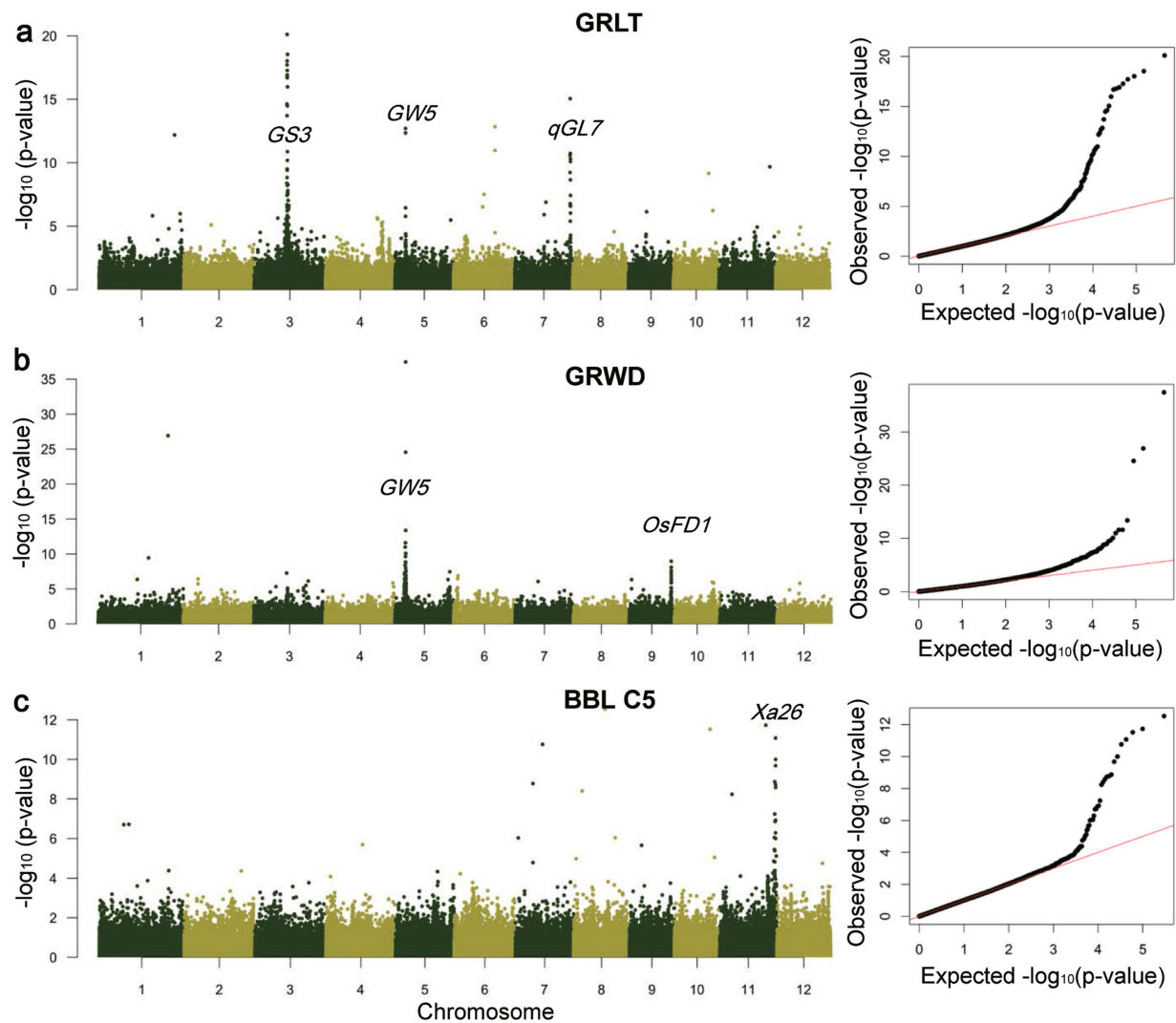
a, b, The numbers of gene families (**a**) and genes (**b**) that emerged at different evolutionary times, from PS1 (single-cell organisms) to PS13 (*O. sativa*). **c**, The age distribution of the core and distributed gene families. **d**, Coding sequence length distribution for Nipponbare RefSeq genes with different ages. **e, f**, SNP variation of the core and distributed genes against the Nipponbare RefSeq. **e**, The density of SNPs in the coding region of core and distributed genes in the 3,010 rice lines; SNP density in core genes is lower than that in candidate core genes (two-sided

Wilcoxon test) and SNP density in candidate core genes is lower than that in distributed genes (two-sided Wilcoxon test). Box plots show the median, box edges represent the first and third quartiles, and the whiskers extend to $1.5 \times$ interquartile range. **f**, K_a/K_s of the core and distributed genes. After removing genes with no synonymous SNPs, there are 3,144 core, 455 candidate core and 800 distributed genes with $K_a/K_s > 1$ and 10,005 core, 1,727 candidate core and 3,957 distributed genes with $K_a/K_s < 1$. Two sided χ^2 -square tests were used to determine the difference of the proportions. ** $P < 0.01$, *** $P < 0.001$.



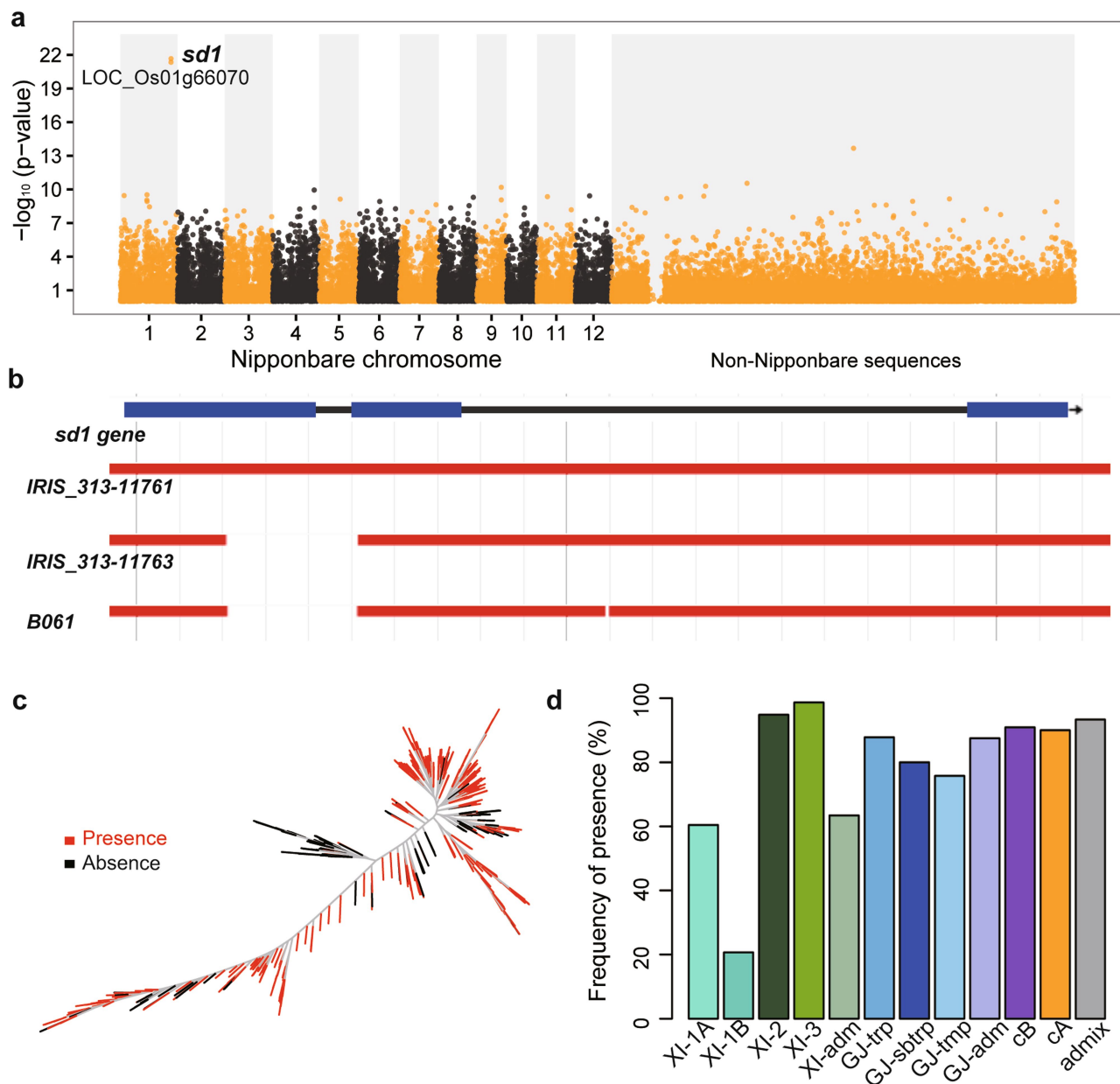
Extended Data Fig. 9 | Haplotypes around the red pericarp (*Rc*) gene. Rows correspond to samples and columns correspond to SNPs. The colour of the rectangle denotes the number of non-Nipponbare alleles in a genotype. **a**, Haplotypes of 3,010 samples in a region ± 25 kb around *Rc* show the presence of many distinctly non-GJ haplotypes that carry

either the wild-type or domesticated allele at the causative deletion site. **b**, Zoomed-in view of a subset of 90 samples and 53 SNPs within the *Rc* gene and 15-kb downstream highlights the wide dispersal of non-GJ domesticated haplotypes.



Extended Data Fig. 10 | Genome-wide association for grain length, grain width and bacterial blight isolate C5. **a–c**, GWAS for grain length (GRLT, $n = 2,012$) (**a**), grain width (GRWD, $n = 2,012$) (**b**) and bacterial blight isolate C5 (BBL C5, $n = 381$) (**c**). GWAS was performed using filtered and linkage disequilibrium-pruned SNPs for historical trait data on source accessions for grain length and grain width (223,743 SNPs)

and for newly acquired lesion length data for bacterial blight isolate C5 (148,999 SNPs). Manhattan plots for linkage disequilibrium-pruned datasets are shown to the left and quartile–quartile plots for expected versus observed $-\log(P)$ values to the right. Major peaks are annotated for known gene loci.



Extended Data Fig. 11 | *sd1* gene and its correlation with plant height.

a, The plot shows the correlation of the gene presence or absence variation with plant height ($n = 323$). The P values were calculated with Spearman's correlation. **b**, Examples show that semi-dwarfism results

from an approximately 385-bp deletion in the *sd1* locus. **c**, Distribution of the presence or absence of the *sd1* gene in the 453 rice accessions. **d**, *sd1* frequencies in rice subpopulations.

A randomized trial of normothermic preservation in liver transplantation

David Nasralla^{1*}, Constantin C. Coussios^{2*}, Hynek Mergental³, M. Zeeshan Akhtar^{1,4}, Andrew J. Butler^{5,20}, Carlo D. L. Ceresa¹, Virginia Chiochia^{6,7}, Susan J. Dutton⁸, Juan Carlos García-Valdecasas⁹, Nigel Heaton¹⁰, Charles Imber¹¹, Wayel Jassem¹⁰, Ina Jochmans^{12,13}, John Karani^{10,14}, Simon R. Knight^{1,15}, Peri Kocabayoglu¹⁶, Massimo Malago¹¹, Darius Mirza³, Peter J. Morris^{1,15}, Arvind Pallan¹⁷, Andreas Paul¹⁶, Mihai Pavel⁹, M. Thamara P. R. Perera³, Jacques Pirenne^{12,13}, Reena Ravikumar¹, Leslie Russell¹⁸, Sara Upponi¹⁹, Chris J. E. Watson^{5,20}, Annemarie Weissenbacher¹, Rutger J. Ploeg¹, Peter J. Friend^{1*} for the Consortium for Organ Preservation in Europe

Liver transplantation is a highly successful treatment, but is severely limited by the shortage in donor organs. However, many potential donor organs cannot be used; this is because sub-optimal livers do not tolerate conventional cold storage and there is no reliable way to assess organ viability preoperatively. Normothermic machine perfusion maintains the liver in a physiological state, avoids cooling and allows recovery and functional testing. Here we show that, in a randomized trial with 220 liver transplantations, compared to conventional static cold storage, normothermic preservation is associated with a 50% lower level of graft injury, measured by hepatocellular enzyme release, despite a 50% lower rate of organ discard and a 54% longer mean preservation time. There was no significant difference in bile duct complications, graft survival or survival of the patient. If translated to clinical practice, these results would have a major impact on liver transplant outcomes and waiting list mortality.

Liver transplantation is the accepted treatment for end-stage liver failure, with one and five year survivals in excess of 90% and 70%, respectively¹. With increasing rates of liver disease², the supply of transplantable organs is no longer able to meet demand. Paradoxically, despite substantial waiting list mortality (for example, 21% in the UK), only 63% of UK deceased donor livers are transplanted¹. Increasing numbers of deceased organ donors in many countries have not been matched by a corresponding rise in the number of transplantable organs. This is mainly because these additional donors tend to be high-risk—either declared dead by cardiovascular criteria (DCD), as opposed to brainstem death donors (DBD), or elderly with multiple co-morbidities (extended criteria donors). Such organs pose a greater risk to the recipient, with a higher probability that the liver will never function (primary non-function (PNF)) or that it will lead to later complications, particularly biliary stricturing.

Despite many advances in liver transplantation, the method of organ preservation has changed very little in almost 30 years³. The liver is flushed and cooled with specialist preservation fluid, then stored in an icebox. This process of static cold storage (SCS) has several limitations. Although SCS slows metabolism by 10- to 12-fold, substantial anaerobic activity continues even at ice temperature⁴. This leads to ATP depletion and accumulation of succinate and other metabolites. These lead to the generation of reactive oxygen species⁵ that are the basis of ischaemia–reperfusion injury, when the organ is re-exposed to oxygenated blood at the time of transplantation. This damage, exacerbated by any prior injury, limits the maximum safe

preservation time of the donor organ. Once cooled, the cessation of normal cellular activity also makes functional assessment impossible.

These shortcomings are particularly problematic in the higher-risk donor organs that form an increasing proportion of current liver transplant practice. The very severe ischaemia–reperfusion-related morbidity that characterizes transplantation of such organs is now a major limitation in meeting the demand for life-saving transplants. To combat the limitations imposed by cold storage, a change in preservation technology is required. In recent years, interest has developed in perfusion at physiological temperature (normothermic machine perfusion (NMP))^{6–9}.

During NMP, the liver is perfused with oxygenated blood, medications and nutrients at normal body temperature to maintain a physiological milieu. Evidence from animal models of both DBD and DCD liver transplantation^{10,11} suggests that this improves the post-transplant survival of transplanted livers, and potentially enables the assessment of organ viability during preservation. The mechanism underlying these improved outcomes is at least partly related to the metabolic resuscitation of the organ that occurs with preservation under physiological conditions. This has been demonstrated through the replenishment of ATP levels¹¹, which in turn contributes to a reduction in the severity of the ischaemia–reperfusion injury that is experienced after transplant^{5,10}.

There is increasing interest in the clinical application of NMP, with several cases described in the recent literature^{6,7}. In 2013, a phase-I study by our group⁹, demonstrated the safety and feasibility of NMP in

¹Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. ²Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. ³Queen Elizabeth Hospital Birmingham, Birmingham, UK. ⁴Target Discovery Institute, University of Oxford, Oxford, UK. ⁵University of Cambridge Department of Surgery, Addenbrooke's Hospital, Cambridge, UK. ⁶Centre for Statistics in Medicine, University of Oxford, Oxford, UK. ⁷Surgical Intervention Trials Unit, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. ⁸Oxford Clinical Trials Research Unit, University of Oxford, Oxford, UK. ⁹Department of Hepatobiliopancreatic and Transplant Surgery, Hospital Clinic, Barcelona, Spain. ¹⁰Institute of Liver Studies, King's College Hospital, London, UK. ¹¹Department of Hepatopancreatobiliary and Liver Transplant Surgery, Royal Free Hospital, London, UK. ¹²Abdominal Transplant Surgery, Department of Surgery, University Hospitals Leuven, Leuven, Belgium. ¹³Laboratory of Abdominal Transplantation, Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium. ¹⁴Department of Radiology, King's College Hospital, London, UK. ¹⁵Centre for Evidence in Transplantation, Clinical Effectiveness Unit, Royal College of Surgeons of England, London, UK. ¹⁶Department of General, Visceral and Transplantation Surgery, University Hospital Essen, Essen, Germany. ¹⁷Department of Radiology, Queen Elizabeth Hospital Birmingham, Birmingham, UK. ¹⁸OrganOx Limited, Oxford, UK. ¹⁹Department of Radiology, Addenbrooke's Hospital, Cambridge, UK. ²⁰National Institute of Health Research (NIHR), Cambridge Biomedical Research Centre, Cambridge, UK. *e-mail: david.nasralla@nds.ox.ac.uk; constantin.coussios@eng.ox.ac.uk; peter.friend@nds.ox.ac.uk

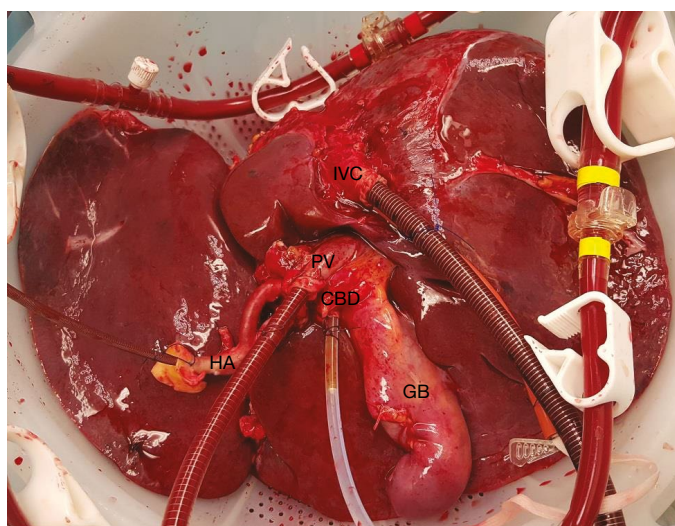


Fig. 1 | Image of liver during normothermic machine perfusion. The hepatic artery (HA), portal vein (PV), inferior vena cava (IVC) and common bile duct (CBD) are all cannulated. The gallbladder (GB) is also present although this was often removed during the retrieval process before NMP. This image has been used with consent from the family of the donor.

20 liver transplant recipients. This was used as the precursor to the present study which, to our knowledge, is the first randomized controlled trial to test the efficacy of machine perfusion against conventional cold storage in liver transplantation.

Livers from adult DBD or DCD donors were eligible for enrolment. Adult patients awaiting a liver-only transplant, excluding those with fulminant liver failure, were eligible. If a suitable liver was allocated to a consented recipient, the liver was randomized to either conventional SCS or NMP. In the SCS arm, the organ retrieval, storage and the transplant were conducted according to standard practice. In the NMP arm, following removal from the donor, the liver was attached to the OrganOx metra NMP device, where it was perfused throughout the duration of preservation (Fig. 1), until the transplanting surgeon was ready to implant it, at which point it was removed from the device. The remainder of the recipient's care followed standard practice.

Daily during the first postoperative week, and at day 10, day 30, month 6 and month 12, biochemical results were recorded as well as graft and patient survival data. At six months, a magnetic resonance imaging scan of the biliary tree (MRCP) was performed to assess evidence of biliary injury. Biological samples were collected and stored in a biobank from each liver and recipient enrolled in the study, for use in further mechanistic studies.

The primary endpoint was defined as the difference between the two treatment arms in the peak level of serum aspartate transaminase (AST) within seven days after transplant. This hepatocellular enzyme is a clinically accepted biomarker, predictive of graft and patient survival¹².

Recruitment

Between 26 June 2014 and 8 March 2016, 334 livers were randomized, with 64 livers subsequently excluded (Fig. 2). Following organ retrieval, a markedly different discard rate between the two trial arms resulted in 100 SCS and 120 NMP livers available for primary outcome reporting, with 101 SCS and 121 NMP livers available for secondary outcome analysis. This discrepancy in group size reduced the study power to 89.7%.

One NMP liver was cold stored due to an accessory left hepatic artery arising from the aorta preventing effective cannulation. Eight NMP livers received machine perfusion for less than four hours (for logistic rather than technical reasons). All of these organs are included in the NMP arm as part of the modified intention to treat analysis. For the per protocol sensitivity analysis, the eight livers perfused for less than four hours were excluded and the single NMP liver that was preserved using SCS was reassigned to the SCS group.

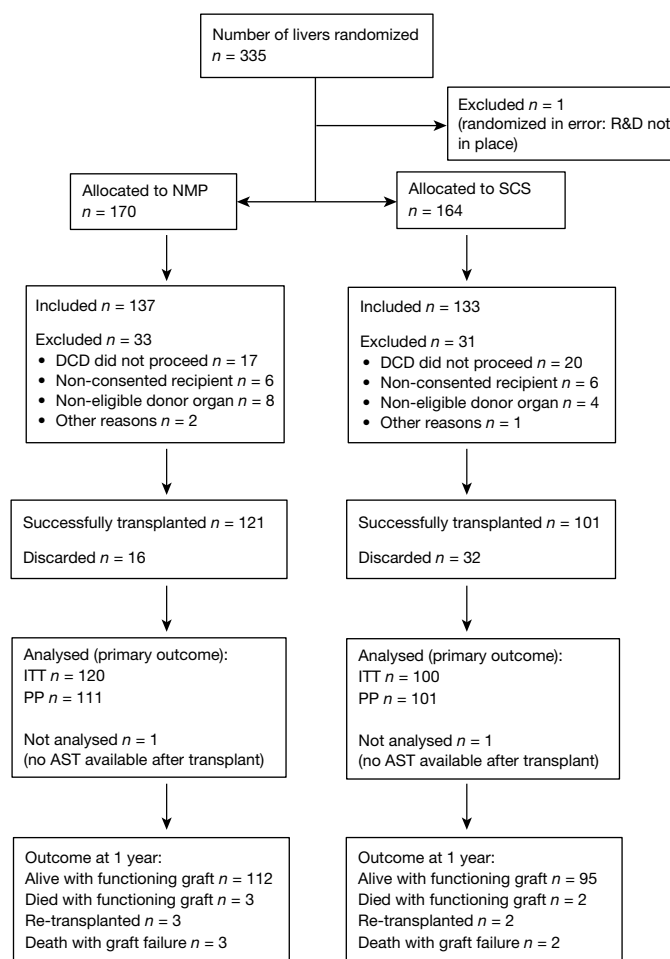


Fig. 2 | CONSORT diagram. CONSORT diagram depicting the outcome for all donor livers enrolled in the trial. ITT, intention to treat; PP, per protocol; R&D, research and development.

Donor, preservation and recipient characteristics

NMP and SCS donor and recipient groups were well-matched (Tables 1, 2). The discard rate was higher in the SCS arm (24.1%; 32 out of 133) than the NMP arm (11.7%; 16 out of 137; Extended Data Table 1). This difference was statistically significant (-12.4% , 95% confidence interval -21.4 to -3.3% ; $P=0.008$). One NMP discard was the result of a device malfunction in an already marginal organ (hepatic artery hypoperfusion due to pinch valve miscalibration; see Supplementary Information).

Functional warm ischaemia time applies only to DCD livers and was measured as the time from the onset of donor hypoxia (oxygen saturation $< 70\%$) or hypoperfusion (systolic blood pressure < 50 mmHg) until the start of cold aortic perfusion in the donor. The median functional warm ischaemia time was longer for NMP than SCS livers (21 min versus 16 min; $P=0.003$).

Total preservation time was measured from the start of cold aortic perfusion in the donor until graft reperfusion in the recipient. The median total preservation time was longer for NMP than SCS livers (11 h 54 min versus 7 h 45 min; $P<0.001$). Within the NMP arm, there was no significant difference in median perfusion time between DBD and DCD livers (9 h 55 min DBD versus 8 h 45 min DCD; $P=0.449$).

Post-reperfusion haemodynamics were documented in 218 cases: post-reperfusion syndrome was more common in the SCS (32 out of 97) than the NMP group (15 out of 121), a statistically significant difference (-20.6% , 95% confidence interval -31.6 to -9.5% ; $P<0.001$). This was despite reduced requirement for vasopressors in NMP livers in the post-reperfusion period (Extended Data Table 2a–c).

Table 1 | Donor demographic details

Stratification factors for all randomized livers	NMP (<i>n</i> = 170)	SCS (<i>n</i> = 164)
Donor type ^a		
DBD	107 (62.9%)	104 (63.4%)
DCD	63 (37.1%)	60 (36.6%)
Donor demographics for all retrieved livers	NMP (<i>n</i> = 137)	SCS (<i>n</i> = 133)
Gender ^a		
Female	54 (39.4%)	57 (42.9%)
Male	81 (59.1%)	76 (57.1%)
Missing	2 (1.5%)	0 (0.0%)
Age ^b	56 (45–67) (16–84)	56 (47–66) (20–86)
Ethnicity ^a		
African–Caribbean	3 (2.2%)	1 (0.8%)
Caucasian	131 (95.6%)	128 (96.2%)
Other	1 (0.7%)	4 (3.0%)
Missing	2 (1.5%)	0 (0.0%)
Cause of death		
CVA	74 (54.0%)	74 (55.6%)
Hypoxia	30 (21.9%)	32 (24.1%)
Trauma	17 (12.4%)	16 (12.0%)
Other	14 (10.2%)	11 (8.3%)
Missing	2 (1.5%)	0 (0.0%)
Body mass index ^b	26.26 (23.66–30.52) (16.42–46.65)	27.01 (23.74–30.56) (17.24–49.96)
Missing	2 (1.5%)	0 (0.0%)
ET–Donor risk index ^b	1.72 (1.47–2.09) (0.98–4.31)	1.72 (1.50–2.10) (1.06–3.49)
Missing	16 (11.7%)	19 (14.3%)

CVA, cerebrovascular accident.

^aFrequency and column percentages are reported.^bMedian, interquartile range (IQR, first brackets) and full range (second brackets) are reported.

Peak AST (primary outcome)

Peak AST during the first 7 days after transplant was reduced by 49.4% in the NMP group compared to SCS when adjusted by centre and donor type (geometric mean ratio 0.506, 95% confidence interval 0.388–0.659; $P < 0.001$). Unadjusted analysis (Student's *t*-test) and sensitivity analysis undertaken in the per-protocol population confirmed these results.

Subgroup analysis showed that the effect of NMP was different in the two donor types (test for interaction $P = 0.012$), although it was statistically significant in both subgroups; the reduction in geometric mean peak AST was greater in DCD (73.3%, 95% confidence interval 53.7–84.6%; $P < 0.001$) than in DBD livers (40.2%, 95% confidence interval 19.3–55.7%; $P = 0.001$). Subgroup analyses for the model for end-stage liver disease (MELD) score and Eurotransplant–donor risk index (ET–DRI) showed no statistically significant differences (data not shown). See Extended Data Table 3a, b and Extended Data Fig. 1 for further analysis. See Table 3 for full outcome results.

Early allograft dysfunction

Data to assess early allograft dysfunction (EAD) rates were available in 216 recipients: the odds of developing EAD in the NMP arm (12 out of 119) were 74% lower than the SCS arm (29 out of 97; odds ratio 0.263, 95% confidence interval 0.126–0.550; $P < 0.001$). A logistic regression model adjusted for donor type, MELD score and ET–DRI showed that the adjusted odds of EAD in the NMP arm were approximately 72% lower than in the cold storage arm (adjusted odds ratio 0.276, 95% confidence interval 0.124–0.611; $P = 0.002$). The difference in EAD rates was partly a result of the difference in peak AST (described above), but also a reflection of differences in bilirubin. The median bilirubin level in the first week postoperatively was lower in NMP recipients (2.25 mg dl⁻¹, 95% confidence interval 1.23–4.28) than in the SCS group (2.87 mg dl⁻¹, 95% confidence interval 1.52–5.00; $P = 0.029$).

Biliary strictures on MRCP

An MRCP was performed on 155 (81 NMP, 74 SCS) of the 222 transplanted trial patients. There was no significant difference in the rate of non-anastomotic strictures for DBD (NMP 7.4% (4 out of 54) versus SCS

Table 2 | Preservation and recipient demographic details

Preservation details for all transplanted livers	NMP (<i>n</i> = 121)	SCS (<i>n</i> = 101)	<i>P</i> value ^a
Functional warm ischaemia time ^b (min) (applies to DCD livers; <i>n</i> = 55 (34 NMP, 21 SCS))	21 (17–25) (9–93)	16 (10–20) (2–32)	0.003
Cold ischaemia time prior to NMP (min) ^c (<i>n</i> = 120)	126 (106.5–143.0) (49–218)	NA	
Machine perfusion time (min) ^c (<i>n</i> = 120)	547.5 (372.5–710.5) (85–1,388)	NA	
Total preservation time from cross-clamp in donor to organ reperfusion in recipient (min)	714 (542–876) (258–1,527)	465 (375–575) (223–967)	0.0000
Steatosis assessed pre-preservation ^{d,e}			0.366
None or mild	91 (75.3%)	89 (88.2%)	
Moderate or severe	29 (24%)	12 (11.9%)	
Missing	1 (0.8%)		
Recipient demographics	NMP (<i>n</i> = 121)	SCS (<i>n</i> = 101)	<i>P</i> value^a
Gender ^d			0.717
Female	35 (28.9%)	27 (26.7%)	
Male	86 (71.1%)	74 (73.3%)	
Donor type ^d			0.209
DBD	87 (71.9%)	80 (79.2%)	
DCD	34 (28.1%)	21 (20.8%)	
Age ^c	55 (48–62) (20–72)	55 (48–62) (22–70)	0.713
Cause of liver failure ^d			0.782
Alcoholic	36 (29.8%)	29 (28.7%)	
Auto-immune hepatitis	2 (1.7%)	5 (5.0%)	
Hepatitis B	3 (2.5%)	2 (2.0%)	
Hepatitis C	4 (3.3%)	4 (4.0%)	
Hepatocellular carcinoma on background of cirrhosis	15 (12.4%)	16 (15.8%)	
Non-alcoholic steato-hepatitis	11 (9.1%)	11 (10.9%)	
Primary biliary cirrhosis	10 (8.3%)	3 (3.0%)	
Primary sclerosis cholangitis	18 (14.9%)	13 (12.9%)	
Other	22 (18.1%)	18 (17.8%)	
Body mass index ^c	26.18 (23.12–32.39) (18.02–50.99)	26.94 (24.36–30.42) (18.91–42.95)	0.626
Missing	0 (0.0%)	1 (1.0%)	
Retransplant ^d	12 (9.9%)	8 (7.9%)	0.605
MELD score ^c (calculated at time of transplant)	13 (10–18) (6–35)	14 (9–18) (6–29)	0.970
UK	13 (10–17) (6–33)	14 (9–18) (6–28)	
Essen, Germany	17 (14–19) (13–23)	15.5 (14–17) (14–17)	
Barcelona, Spain	16 (8–26) (8–35)	14 (9–16) (8–29)	
Leuven, Belgium	19 (13.5–25.0) (13–26)	16 (16–20) (9–27)	
eGFR ^c	87.36 (69.61–107.66) (33.45–156.43)	92.22 (69.72–104.24) (30.19–155.04)	0.928
Missing	4 (3.3%)	3 (3.0%)	
ET–Donor risk index ^c	1.70 (1.47–2.07) (0.98–4.31)	1.71 (1.50–2.01) (1.06–3.49)	0.610
Missing	13 (10.7%)	13 (12.9%)	

NA, not applicable. eGFR, estimated glomerular filtration rate.

^a χ^2 tests and non-parametric Mann–Whitney *U*-tests were used for categorical and continuous variables, respectively. No adjustment for multiple comparisons were made.^bFunctional warm ischaemia applies to DCD donors and is measured from the onset of functional warm ischaemia (systolic blood pressure < 50 mmHg or O₂ saturation < 70%) to cross-clamp.^cMedian, IQR and full range are reported.^dFrequency and column percentages are reported.^eMeasurement of the degree of steatosis was based on clinical assessment by the retrieval surgeon.

5.4% (3 out of 55); $P = 0.678$) or DCD (NMP 11.1% (3 out of 27) versus SCS 26.3% (5 out of 19); $P = 0.180$) livers. Only one patient in each trial arm developed clinically relevant evidence of ischaemic cholangiopathy in the first year after transplant, both of whom were re-transplanted.

Table 3 | Trial outcomes

	NMP (<i>n</i> = 121) ^a	SCS (<i>n</i> = 101) ^a	Effect (95% CI) ^b	<i>P</i> value
Peak AST				
ITT ^c				
Adjusted	488.1 (408.9–582.8)	964.9 (794.5–1,172.0)	0.5 (0.4–0.7)	0.0000
Unadjusted	484.5 (406.4–577.6)	973.7 (795.2–1,192.3)	0.5 (0.4–0.6)	0.0000
Test for interaction by donor type				0.012
Subgroup analysis by donor type				
DBD	526.2 (427.3–647.9)	880.2 (708.5–1,093.5)	40.2% (19.3–55.7%)	0.0009
DCD	389.7 (278.0–546.4)	1,458.1 (944.7–2,250.5)	73.3% (53.7–84.6%)	0.0000
PP analysis	498.6 (414.8–599.4)	982.9 (810.4–1,192.2)	0.5 (0.4–0.7)	0.0000
Secondary outcomes				
Discard rates ^d	16 (11.7%)	32 (24.1%)	–12.4% (–21.4 to –3.3%)	0.008
Primary non-function ^e	1 (0.8%)	0 (0.0%)	NA	NA
Post-reperfusion syndrome	15 (12.4%)	32 (33.0%)	–20.6% (–31.6 to –9.6%)	0.0002
Post-reperfusion lactate ^f	3.6 (2.6–4.2)	4.1 (3.2–5.0)		0.018
Early allograft dysfunction	12 (10.1%)	29 (29.9%)	0.263 (0.126–0.550)	0.0002
Biochemical liver tests^f (average value over day 1–7)				
Bilirubin (μmol l ^{–1})				
Days 1–7	38.5 (21.0–73.2)	49.1 (26.0–85.5)		0.029
30 days	13.0 (8.0–22.1)	13.0 (9.1–21.0)		0.479
6 months	9.1 (6.0–15.1)	9.1 (6.0–13.0)		0.671
AST (IU l ^{–1})				
Days 1–7	167.5 (98.0–320.7)	318.5 (152–611.5)		0.0000
30 days	20 (14–35)	22 (15–40)		0.707
6 months	23 (18–33)	23 (18–37)		0.931
γGT (IU l ^{–1})				
Days 1–7	268.1 (156.3–408.3)	301 (201.1–443.9)		0.157
30 days	178 (109.5–410.0)	200 (96.0–397.5)		0.949
6 months	47 (28–144)	47 (26–128)		0.452
INR				
Days 1–7	1.2 (1.2–1.4)	1.2 (1.2–1.4)		0.644
30 days	1.1 (1.0–1.2)	1.1 (1.0–1.2)		0.735
6 months	1.1 (1.0–1.2)	1.1 (1.0–1.1)		0.167
Creatinine (μmol l ^{–1})				
Days 1–7	92.8 (60.1–121.1)	97.2 (67.2–143.2)		0.139
30 days	82.2 (66.3–104.3)	90.2 (72.5–121.1)		0.019
6 months	99.9 (81.3–117.6)	99.9 (83.1–134.4)		0.265
Lactate (mmol l ^{–1})				
Day 1–7	1.3 (1.0–1.7)	1.1 (0.9–1.6)		0.130
Other outcomes				
Need for RRT (number (percentage) of patients)				
Day 1–7 after transplant	26 (21.5%)	19 (18.8%)	2.7% (–7.9 to 13.2%)	0.621
30 days	27 (22.3%)	20 (19.8%)	2.5 (–8.2 to 13.3%)	0.648
6 months	27 (22.3%)	21 (20.8%)	1.5% (–9.3 to 12.4%)	0.784
Duration of RRT day 1–7 ^f	4 (2–6)	5 (4–6)		0.346
Length of hospital stay ^f	15 (10–24)	15 (11–24)		0.926
Length of ICU stay ^f	4 (2–7)	4 (3–7)		0.339
Graft survival at 1 year	0.950 (0.893–0.977)	0.960 (0.897–0.985)		0.707
Patient survival at 1 year	0.958 (0.902–0.982)	0.970 (0.909–0.990)		0.671

CI, confidence interval.

^aTotal number of livers transplanted and analysed overall. Primary outcome analysed on *n* = 220 due to unavailability of AST values during the first seven days after transplant. Specific outcomes may have different denominators due to some missing data.^bEffect reported is: Percentage reduction (from geometric mean ratio) for peak AST; odds ratio for early allograft dysfunction; difference in proportions (%) for discard rates, post reperfusion syndrome and need for renal replacement therapy (RRT); not reported for outcomes for which medians are reported, for survival scores and for tests for interactions of subgroup analysis (only *P* values are reported).^cIntention to treat (ITT) analysis was adjusted for donor type and transplant centre.^dDenominators for the discard rates is the total number of livers retrieved (*n* = 270 (NMP, *n* = 137; SCS *n* = 133)).^eTest not performed due to few events and no events in one arm.^fMedian and IQR are reported, a non-parametric Mann–Whitney *U*-test was used.

Similarly, there was no significant difference in the rate of anastomotic strictures for DBD (NMP 40.7% (22 out of 54) versus SCS 41.8% (23 out of 55); *P* = 0.909) or DCD (NMP 48.1% (13 out of 27) versus SCS 57.9% (11 out of 19); *P* = 0.515) livers.

Hospital stay, graft and patient survival

There was no difference in median intensive care unit (ICU) stay (4 days NMP versus 4 days SCS; *P* = 0.339), hospital stay (15 days NMP versus 15 days SCS; *P* = 0.926) or the need for renal replacement therapy in the first postoperative week (2.7%, 95% confidence interval –7.9 to 13.2%; *P* = 0.621).

One NMP liver developed PNF (see Supplementary Information). There were no PNF cases in the SCS arm. Overall 10 recipients died

during follow-up, producing a one-year survival of 0.949 (95% confidence interval 0.890–0.977) in the NMP group and 0.958 (95% confidence interval 0.902–0.982) in the SCS group (*P* = 0.901). Two deaths in the SCS group and three deaths in the NMP group were due to graft failure.

Graft survival at one year was 0.950 (95% confidence interval 0.893–0.977) and 0.960 (95% confidence interval 0.897–0.985) in the NMP and SCS groups, respectively (*P* = 0.695). The causes of graft failure in the SCS arm were hepatic artery thrombosis (*n* = 3) and ischaemic cholangiopathy (*n* = 1). The causes of graft failure in the NMP arm were hepatic artery thrombosis (*n* = 2), ischaemic cholangiopathy (*n* = 1), non-thrombotic infarction (*n* = 1), inferior vena cava occlusion (*n* = 1) and PNF (*n* = 1). (Extended Data Fig. 2a, b).

For more detailed analysis of trial outcomes please see Supplementary Information.

Perfusion characteristics indicative of organ quality

The following continuously monitored parameters (mean \pm s.d.) by the third hour of NMP were measured for all livers that went on to be successfully transplanted (Extended Data Figs. 3, 4). The measured haemodynamic parameters were: hepatic artery flow (280 ± 120 ml min⁻¹) and portal vein flow (1.11 ± 0.21 min⁻¹). The measured metabolic parameters were: pH (7.31 ± 0.17) and lactate clearance from 9.99 ± 3.13 mmol l⁻¹ at 15 min NMP to 0.93 ± 0.63 mmol l⁻¹ by 4 h NMP. The measured synthetic parameter consisted of bile production (9.17 ± 11.16 ml h⁻¹). Notably, 18 transplanted NMP livers produced no/minimal bile during perfusion. All but one of these functioned after transplant. There was no correlation between bile production and post-transplant liver function or later development of non-anastomotic biliary strictures.

One NMP liver developed PNF. This liver was persistently acidotic with lactate > 4 mmol for the duration of NMP. No other liver with these characteristics was transplanted.

Following transplant, 28 livers displayed minimal preservation injury (MPI; peak AST < 250 IU l⁻¹) and 25 showed evidence of severe preservation injury (SPI; peak AST $> 1,000$ IU l⁻¹). The donors in these groups were well-matched in all characteristics other than sex (Extended Data Table 4). During NMP, there was a difference in baseline perfusate alanine aminotransferase (ALT) (MPI 171 IU l⁻¹ versus SPI 669 IU l⁻¹; $P = 0.005$) and lactate dehydrogenase (LDH) (MPI 1,073 IU l⁻¹ versus SPI 1,838 IU l⁻¹; $P = 0.01$) between the two groups. Levels of these enzymes, as well as γ -glutamyltransferase (γ GT), increased more rapidly during the first 8 h of NMP in the SPI group (ALT, an increase of 56 IU l⁻¹ versus an increase of 461 IU l⁻¹, $P < 0.001$; LDH, an increase of 483 IU l⁻¹ versus an increase of 980 IU l⁻¹, $P = 0.06$; γ GT, an increase of 23 IU l⁻¹ versus increase of 104 IU l⁻¹, $P = 0.004$). MPI livers showed a reduction in measurable levels of haemolysis (haemolysis index) as NMP progressed, in contrast to SPI livers in which the levels of haemolysis rose (MPI, a decrease of 0.04 U versus SPI, an increase of 0.09 U; $P = 0.03$). Bile production was greater in the MPI group (MPI 13.1 ml h⁻¹ versus SPI 7.8 ml h⁻¹; $P = 0.03$). Lactate clearance was similar in each group. Post-reperfusion syndrome was less common in the MPI group (MPI 0% (0 out of 28) versus SPI 24% (6 out of 25); $P = 0.007$). One NMP liver with perfusate transaminases in excess of 20,000 IU l⁻¹ was transplanted successfully.

Adverse events

The proportion of patients for whom adverse events were reported (Extended Data Tables 5a–c, 6) was similar in the two arms (55.4% NMP, 95% confidence interval 46.1–64.4% versus 57.4% SCS, 95% confidence interval, 47.2–67.2%) with a larger total number of events reported for SCS livers (128 NMP versus 164 SCS). Of these, a greater proportion of the serious adverse events (Clavien–Dindo grade \geq IIIb) were in the SCS than NMP arm (16.4% NMP (21 out of 128) versus 22% SCS (36 out of 164)). No statistical tests were applied to these data.

Discussion

To our knowledge, this is the first randomized controlled trial to compare any type of machine perfusion technology with conventional static cold storage in human liver transplantation.

The trial demonstrated significant reductions in peak AST and EAD rates in NMP livers; this is of clinical relevance as both are clinically accepted biomarkers for long-term graft and patient survival^{12,13}. These benefits are consistent with previous animal work¹⁰ and the phase-I clinical study⁹ that preceded this trial, both of which showed post-transplant AST reductions in NMP livers. No differences were seen in graft or patient survival: a much larger trial is required to test this outcome. It is notable that these reductions in peak AST and EAD rates were achieved in the context of improved organ utilization and longer preservation times, both of which have implications in terms

of addressing the donor shortage and logistical barriers that currently limit liver transplants.

DCD donors represent a largely untapped source of organs, comprising 42% of UK deceased donors, but only 21% of transplanted livers¹⁴. Utilization of DCD livers is limited by poorer outcomes (PNF and ischaemic cholangiopathy) compared with DBD livers. Allowing the limitations of small group analyses, in this study NMP DCD liver primary outcome data were superior to those of both DCD and DBD livers preserved using SCS. In fact, the primary outcome of DCD NMP livers was superior to that of DBD livers preserved by NMP: this was possibly owing to a selection bias, both of donors (lower threshold to decline DCD donors) and recipients (fitter patients selected for higher-risk organs). The AST differences are in the context of longer functional warm ischaemia times, longer preservation times and fewer organ discards in the NMP arm, suggesting that NMP may be achieving the desired objective of increasing organ utilization without compromising outcome. If these findings were translated into clinical practice, the increase in organ utilization would have substantial implications for waiting list mortality, which is currently approximately one in five patients¹.

The longer preservation times in the NMP group were not planned, but were all within the maximum perfusion time defined in the protocol. There was no stipulation in the trial protocol that the preservation times should be matched. As clinicians gained experience, it appeared that some centres had started to organize their operating schedule according to the preservation method, although no overall difference between arms was seen in the proportion of transplants occurring in daylight hours. If, as appears to be the case, NMP can safely extend preservation times without compromising outcomes, this will have implications for operating department planning as well as organ utilization.

There were over 50% fewer discarded organs in the NMP group, resulting in 20% more transplanted livers (121 NMP versus 101 SCS). The SCS discard rate of 23.7% was higher than the 17% reported in UK registry data¹⁴, and may reflect the high proportion of DCD livers enrolled in the trial; the discard rate of retrieved DCD livers in the UK is 30%¹⁴. This reported difference in organ utilization is likely to be an underestimation of the full potential impact that NMP could have on transplant numbers. The trial stipulated that only livers considered transplantable according to standard practice could be enrolled. For the full extent of improved organ utilization to be measured, livers would need to be randomized to NMP or SCS before being offered for transplant; this should form the basis of a future study. An increase of 20% or more in the number of transplantable donor livers would have a transformative effect on the mortality on liver transplant waiting lists around the world.

The haemodynamic characteristics of the NMP recipients following reperfusion were measurably superior to those of SCS recipients, in line with previously reported findings¹⁵. This did not translate into a difference in ICU stay, hospital stay or need for renal replacement therapy between the two groups, despite previous reports showing a correlation between peak AST and renal replacement therapy¹⁶. The magnitude of the reperfusion syndrome is a factor in determining the eligibility of the sickest patients for high-risk organs, due to the limited capacity of such patients to tolerate cardiovascular instability; NMP might therefore increase the options for the most urgent patients.

Perhaps the greatest limitation to more widespread utilization of DCD livers is the high rate of clinically important non-anastomotic biliary strictures (NAS) which lead to a high rate of graft failure; this is believed to develop due to the vulnerability of the biliary tree to prolonged warm ischaemia. The rate of NAS in the NMP DCD group (11.1%) was lower than in SCS (26.3%) livers, despite longer functional warm ischaemia times. This did not reach statistical significance, which may be a function of sample size; the trial was not powered for this outcome. Reported rates of NAS in DCD transplants vary from 10 to 30%^{17,18}, but these are in patients with symptoms suggesting biliary pathology, rather than those only apparent on imaging; biliary

investigations are usually only performed for clinical indications (typically deranged liver function). Prior to this study, the radiological incidence of both anastomotic and non-anastomotic strictures in asymptomatic patients was unknown; in particular, there is no real benchmark against which to compare the rate of NAS seen in the DCD SCS group. Apart from the two patients retransplanted for ischaemic cholangiopathy, almost all of the remaining patients with radiological evidence of NAS had normal liver function at one year; this questions the clinical relevance of a protocol MRCP at six months. The longer-term follow-up of these patients will shed light on the importance of a radiological diagnosis of biliary stricturing in patients with normal graft function, and the role of MRCP as an endpoint in future trials.

As well as demonstrating improved graft preservation, this trial tested the feasibility, usability and safety of NMP, a vital component of the evaluation of any new technology. It showed that the logistical challenges of NMP can be met successfully within clinical practice. Over 120 NMP livers were transplanted in seven transplant centres across four European countries. Nonetheless, adoption of this technology into clinical practice may necessitate changes in the organ retrieval process, particularly with respect to technical support and transport arrangements. It remains to be seen whether NMP is required for the full duration of an organ's preservation or can equally well be applied after a short period of SCS when the organ reaches the transplanting centre—this would simplify the logistics but may not be suitable for the most marginal organs¹⁹. A phase-II study to test this has recently completed enrolment in the UK (NCT03176433).

For this new technology to be supported by healthcare funders, a health-economic case is needed. The results of this study suggest that benefits will accrue not only from improved early graft function and transplantation logistics, but also from improved utilization. Secondary economic benefits will accrue from logistic changes, enabling transplants to be moved predominantly into daytime operating, with reduction in staffing costs and likely improvements in outcome. More timely intervention will also bring economic benefits—earlier transplantation is associated with lower morbidity and cost.

The effects of NMP demonstrated in this study are unequivocal with respect to the primary endpoint, implying a benefit in livers currently used for transplantation. However, the greatest benefit may be realized by applying this technology to livers outside current acceptance criteria, in order to transplant organs currently deemed untransplantable. Algorithms to assess organ viability, based on data obtained during NMP, will be essential if this potential is to be realized¹⁰. This study sheds some light on which perfusion parameters may be used to assess organ quality: bile production, acid–base stability, lactate clearance, perfusate transaminase levels, falling measurable haemolysis—all correlate with the degree of preservation injury evident after transplant. However, all but one of the livers transplanted in the NMP group functioned postoperatively, including one NMP liver with perfusate transaminases in excess of 20,000 IU l⁻¹ and 18 livers with minimal bile production. Data from much larger numbers of NMP transplants (typically from a registry) would be required to determine specific markers of viability.

The importance of bile production during NMP is unclear. Preliminary evidence from our group²⁰ suggests that preservation injury causes impaired hepatocellular uptake of bile salts. We have shown evidence of progressive accumulation of bile salts in the perfusate of livers with high post-transplant transaminase levels; something that also correlates with poor bile production during NMP. The extent and nature of the injury required to produce this effect is not clear but does appear to reflect organ quality rather than viability.

High-risk organs (for example, those with steatosis) may benefit from therapeutic interventions delivered during NMP: several groups are exploring potential strategies, including stem cell treatments, de-fating agents and immunological modification of the organ. Future trials may be needed to formally test the size of the effect of NMP on organ utilization; for this it will be necessary to randomize livers at the time of organ offering rather than the time of retrieval. Organ utilization, or

organ utilization with 12-month graft survival (functional utilization) would be a logical primary endpoint for a study of this sort.

This study describes the formal clinical evaluation of a novel technology in liver transplantation, and could herald the start of a new era of intervention during organ preservation. It represents a first, necessary step in demonstrating that NMP is feasible, safe and effective in clinical practice; the fact that the study has definitively met its primary endpoint should now enable the exploration of the technology's wider potential.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0047-9>.

Received: 28 September 2017; Accepted: 8 March 2018;

Published online 18 April 2018.

1. *Annual Report on Liver Transplantation 2016/2017* (NHS Blood and Transplant, 2017).
2. APPHG. Liver disease: today's complacency, tomorrow's catastrophe. The All-Party Parliamentary Hepatology Group (APPHG) inquiry into improving outcomes in liver disease. (March, 2014).
3. Todo, S. et al. Extended preservation of human liver grafts with UW solution. *J. Am. Med. Assoc.* **261**, 711–714 (1989).
4. Clavien, P. A., Harvey, P. R. & Strasberg, S. M. Preservation and reperfusion injuries in liver allografts. An overview and synthesis of current studies. *Transplantation* **53**, 957–978 (1992).
5. Chouchani, E. T. et al. Ischaemic accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature* **515**, 431–435 (2014).
6. Watson, C. J. et al. Preimplant normothermic liver perfusion of a suboptimal liver donated after circulatory death. *Am. J. Transplant.* **16**, 353–357 (2016).
7. Mergental, H. et al. Transplantation of declined liver allografts following normothermic ex-situ evaluation. *Am. J. Transplant.* **16**, 3235–3245 (2016).
8. Perera, T. et al. First human liver transplantation using a marginal allograft resuscitated by normothermic machine perfusion. *Liver Transpl.* **22**, 120–124 (2016).
9. Ravikumar, R. et al. Liver transplantation after ex vivo normothermic machine preservation: a phase 1 (first-in-man) clinical trial. *Am. J. Transplant.* **16**, 1779–1787 (2016).
10. Brockmann, J. et al. Normothermic perfusion: a new paradigm for organ preservation. *Ann. Surg.* **250**, 1–6 (2009).
11. Xu, H. et al. Excorporeal normothermic machine perfusion resuscitates pig DCD livers with extended warm ischemia. *J. Surg. Res.* **173**, e83–e88 (2012).
12. Eisenbach, C. et al. An early increase in gamma glutamyltranspeptidase and low aspartate aminotransferase peak values are associated with superior outcomes after orthotopic liver transplantation. *Transplant. Proc.* **41**, 1727–1730 (2009).
13. Olthoff, K. M. et al. Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl.* **16**, 943–949 (2010).
14. *Organ Donation and Transplantation Activity Report 2016/17* (NHS Blood and Transplant, 2017).
15. Angelico, R. et al. Normothermic machine perfusion of deceased donor liver grafts is associated with improved postreperfusion hemodynamics. *Transplant. Direct* **2**, e97 (2016).
16. Leithead, J. A. et al. Hepatic ischemia reperfusion injury is associated with acute kidney injury following donation after brain death liver transplantation. *Transpl. Int.* **26**, 1116–1125 (2013).
17. Jay, C. L. et al. Ischemic cholangiopathy after controlled donation after cardiac death liver transplantation: a meta-analysis. *Ann. Surg.* **253**, 259–264 (2011).
18. Mourad, M. M., Algarni, A., Liossis, C. & Bramhall, S. R. Aetiology and risk factors of ischaemic cholangiopathy after liver transplantation. *World J. Gastroenterol.* **20**, 6159–6169 (2014).
19. Reddy, S. et al. Non-heart-beating donor porcine livers: the adverse effect of cooling. *Liver Transpl.* **11**, 35–38 (2005).
20. Abstracts of the 18th Congress of the European Society for Organ Transplantation, 24–27 September 2017, Barcelona, Spain. *Transpl. Int.* **30** (Suppl 2), 5–576 (2017).

Acknowledgements This study was performed by the Consortium for Organ Preservation in Europe (COPE). We thank the European Commission for their support through the Seventh Framework Programme. The following organisations, groups and individuals also made substantial contributions without which this trial could not have been completed successfully: NHS Blood and Transplant; the Surgical Intervention Trials Unit, University of Oxford; the Clinical Trials and Research Governance unit, University of Oxford; Centre for Evidence in Transplantation, Royal College of Surgeons of England; the Liver Transplant Coordinators, anaesthetists and liver unit physicians at the Queen Elizabeth Hospital, Birmingham, Addenbrooke's Hospital, Cambridge, King's College Hospital, London, the Royal Free Hospital London, Hospital Clinic,

Barcelona, University Hospitals, Leuven, University Hospital, Essen; M. Soo, S. Morrish, C. Morris, L. Randle, R. Macedo Arantes, R. Morovat, A. Elsharkawy, G. Hirschfield, P. Muiesan, J. Isaac, J. Grayer, B. Buchholz, H. Vilca-Melendez, A. Zamalloa, D. Chasiotis, S. Khorsandi, B. Davidson, D. Sharma, A. Esson, D. Monbaliu, S. Mertens, S. Swoboda, J. Neuhaus, T. Benkő, V. Molina, R. Kumar, A. Bradley, M. Laspeyres, B. Patel, A. Mukwamba, S. Banks, the COPE Transplant Technicians, the Specialist Nurses in Organ Donation and, of course, all of the donors and their families. This study was funded by a European Commission Seventh Framework Programme (FP7) Grant (No 305934).

Reviewer information Nature thanks S. Schneeberger, S. G. Tullius and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.N., S.R.K., R.J.P., C.C.C. and P.J.F. designed this study with help from other authors. R.J.P. is Coordinator of the COPE Consortium. M.Z.A. and R.J.P. oversaw the collection of samples and establishment of the trial biobank. D.N., C.D.L.C., A.W., H.M., M.T.P.R.P., D.M., W.J., N.H., C.I., M.M., R.R., A.J.B., C.J.E.W., I.J., J.P., P.K., A.Pau., M.P. and J.C.G.-V. were responsible for the clinical conduct of the study at the respective trial sites. D.N., S.R.K., V.C. and S.J.D. were responsible for statistical design and analysis. L.R. and C.C.C. provided device support and expertise to all trial sites. D.N., S.U., A.Pal. and

J.K. were responsible for MRCP image analysis. P.M., S.R.K. and S.J.D. provided governance oversight to ensure the study adhered to all regulatory and ethical requirements. All authors reviewed the manuscript.

Competing interests P.J.F. is a co-founder, chief medical officer and consultant to OrganOx Limited and also holds shares in the company. C.C.C. is a co-founder, chief technical officer and consultant to OrganOx Limited and also holds shares in the company. Neither P.J.F. nor C.C.C. were involved in the selection, recruitment or transplantation of patients in this study.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0047-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0047-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.N. or C.C.C. or P.J.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Study design. This investigator-led, multinational, open-label, two-arm, parallel randomized controlled trial included seven liver transplant centres from the UK (Addenbrooke's Hospital, Cambridge; King's College Hospital, London; Queen Elizabeth Hospital, Birmingham; Royal Free Hospital, London), Belgium (Universitaire Ziekenhuizen, Leuven), Spain (Hospital Clínic de Barcelona, Barcelona) and Germany (Universitätsklinikum, Essen), and was part of the EU-funded Consortium for Organ Preservation in Europe (COPE, <http://www.cope-eu.org/>). Approval was obtained from national research ethics committees and medical device regulatory bodies in each trial region, in particular the London–Dulwich National Research Ethics Committee (NREC) and the Medicines and Healthcare Regulatory Agency (MHRA) in the UK; the Federala Agentschap voor Geneesmiddelen en Gezondheidsproducten (FAGG) and the Commissie Medische Ethiek of Universitaire Ziekenhuizen, Leuven, Belgium; the Comité Ético de Investigación Clínica of the Hospital Clínic de Barcelona; the Deutsche Ärztekammer and the ethics committee of University Hospital Essen, Germany. The trial protocol was registered before recruitment (ISRCTN 39731134). All relevant ethical regulations relating to the conduct of this study were followed at each trial site. The trial is reported in accordance with the CONSORT statement²¹.

No major amendments were made to the trial design after the start of recruitment.

Eligibility and consent. Inclusion criteria for donors and recipients were deliberately broad to represent the full spectrum of clinical practice. Whole livers from DBD and DCD (Maastricht category III²²) donors at least 16 years of age were eligible. Specific donor consent was not required for trial inclusion. No organs were procured from prisoners. Recipients were eligible provided they were at least 18 years old and listed for a liver-only transplant, excluding those with fulminant liver failure, owing to the poor prognosis of this group regardless of organ quality. Potential participants were consented while on the waiting list; consent was affirmed on the day of transplantation. The consent included the recording of anonymized data for trial purposes and the collection of biological samples for storage in the trial biobank (see 'Sample collection'). No patient identifiable data were collected.

Randomization. Once an eligible donor organ was allocated to a consented recipient and the availability of the NMP device and team was confirmed, the liver was randomized. All clinical decisions thereafter, including graft suitability and procedure scheduling, were made independently of the trial team.

Using an online randomization tool, livers were assigned to NMP or SCS with 1:1 allocation ratio as per a computer-generated randomization schedule, using variable block size, stratified by transplant centre and donor type (DBD/DCD). The unit of randomization was donor livers rather than recipients, but analysis is reported for the transplant recipients.

Static cold storage group. Livers randomized to SCS were retrieved, preserved, transported and transplanted according to local standard practice.

Normothermic machine perfusion group. The OrganOx metra normothermic liver perfusion device was used (Extended Data Fig. 5a), which enables automated organ preservation for up to 24 h. Following randomization to NMP, the device and accompanying researcher were transported to the donor hospital. The device was set-up during the retrieval procedure, as has been previously described⁹. A sterile disposable set was installed on to the device and primed with 500 ml gelofusine (B. Braun Ltd) and three units of donor-matched packed red blood cells. Antibiotics were given at the outset and heparin, insulin, prostacyclin, bile salts and fat-free parenteral nutrition were infused during the perfusion (Extended Data Fig. 5b).

Following retrieval of the donor organ, and while still at the donor hospital, the liver back-table operation was performed²³, followed by cannulation of the hepatic artery, portal vein, inferior vena cava and bile duct. The liver was connected to the NMP device and perfusion commenced (Fig. 1). During the early part of the perfusion sodium bicarbonate was added incrementally to achieve a physiological pH. The OrganOx metra perfusion device incorporates online blood gas measurement (Terumo CD1-500) together with software-controlled algorithms to control P_{O_2} and P_{CO_2} (within physiological limits), temperature (37°C), mean arterial pressure (65–75 mm Hg) and, inferior vena cava pressure (0–2 mm Hg). Typical blood flows of 200–400 ml min⁻¹ (artery) and 1,000–1,200 ml min⁻¹ (portal vein) were obtained. Glucose was measured manually and the value entered into the device. If glucose fell below 10 mmol l⁻¹ this automatically triggered the infusion of a fat-free TPN mixture (Nutriflex Special, B. Braun Ltd) into the perfusate.

NMP continued throughout the duration of transport and storage until the transplanting team were ready to implant the liver. The minimum protocol-stipulated NMP duration was 4 h, the time needed for ATP repletion in animal studies¹¹. The maximum allowed NMP duration was 24 h in line with the experience in the phase-I study and the regulatory approval for the device⁹.

Sample collection. Tissue biopsies (donor liver and bile duct), recipient blood and urine were collected at pre-specified time points from every liver/transplanted patient in the study. In addition to these, samples of perfusate fluid and bile were

collected from every NMP liver. These were stored in a central biobank established by the COPE Consortium for use in ongoing mechanistic studies. Each sample was allocated a unique bar code, which the biobank coordinator was able to match to a specific trial identification number. No patient identifiable data were associated with each sample.

Study end points. The primary endpoint was defined as the difference between the two treatment arms in the peak level of serum AST within seven days after transplant. This is a clinically accepted biomarker, predictive of primary non-function as well as graft and patient survival^{12,24} and is also associated with histological evidence of moderate to severe perfusion injury^{25,26}.

A surrogate marker of graft survival was used in this trial for two reasons: (1) the relatively high survival rates in liver transplantation (> 90%) and (2) the multifactorial causes of graft loss. A trial based directly on graft or patient survival would have had to be unfeasibly large.

In order to ensure consistency and to minimise the hypothetical AST 'wash-out' effect in the NMP-treated organs, the first post-transplantation value was measured between 12 and 24 h after reperfusion.

Secondary end points included: (1) organ discard rate (after retrieval); (2) post-reperfusion syndrome²⁷: > 30% drop in mean arterial pressure persisting for > 1 min within five minutes of reperfusion; (3) primary non-function: irreversible graft dysfunction, for non-technical and non-immunological causes, leading to death or emergency liver replacement during the first 10 days after liver transplantation; (4) early allograft dysfunction¹³ as indicated by any one of the following clinical indicators: (i) bilirubin > 170 µmol l⁻¹ on day 7 after transplant; (ii) INR > 1.6 on day 7 after transplant; (iii) peak-AST > 2,000 IU l⁻¹ during the first 7 days; (4) length of hospital and ICU stays; (5) need for renal replacement therapy; (6) evidence of cholangiopathy on MRCP at six months; (7) graft and patient survival at one year.

Full details of all secondary outcomes are available in the trial protocol²⁸.

Six-month MRCP. An MRCP scan was performed six months (range 5–7 months) after transplant to evaluate the biliary tree for features of cholangiopathy evident by biliary strictures. All scans were reviewed by two independent radiologists blinded to the method of organ preservation with disparities adjudicated by a third radiologist. Owing to the lack of any existing grading system for biliary strictures, a system was agreed to in advance by consensus among the radiologists to allow definitive categorization of the presence and site of strictures. The findings were reported as follows: (1) normal biliary tree; (2) anastomotic stricture (> 70% of luminal diameter); (3) unequivocal evidence of non-anastomotic stricture anywhere in the biliary tree; (4) both anastomotic and non-anastomotic biliary strictures.

Statistical analysis. Previous data from Universitaetsklinikum Essen, Germany (A. Pau. & S.R.K., unpublished observations), demonstrated the geometric mean of peak AST to be 608.59 IU l⁻¹ in patients transplanted following SCS. The present study was powered to detect a (clinically relevant) 33% reduction in peak AST with 90% power at a 5% significance level, requiring 220 transplanted livers (110 per arm).

Results are reported as a modified intention-to-treat analysis. A per-protocol sensitivity analysis was also performed excluding livers that received machine perfusion outside the protocol specified range (4–24 h) and comparing the groups according to the treatment actually received. Livers randomized but not retrieved were excluded from the analysis.

Primary outcome was analysed using ANOVA with adjustment for stratification factors. The peak AST was calculated for each recipient with at least two values available. Missing AST values were not imputed. Binary outcomes were assessed using test for proportions or logistic regression to adjust for potential confounders and report odds ratios. Continuous outcomes were compared using a Student's *t*-test, if normally distributed, or by Mann–Whitney *U*-test otherwise. Time-to-event outcomes were analysed using Kaplan–Meier estimates and log-rank tests. Outcomes are reported with 95% confidence intervals and *P* values to three decimal places. *P* < 0.05 was regarded as statistically significant.

Pre-specified subgroup analyses were performed for donor type (DCD versus DBD), donor risk index (ET-DRI) and MELD score using tests for interaction and reported using forest plots. Interaction methods were used to look for consistency of treatment effect across the different subgroups and reported using forest plots. The study was not powered to detect differences in the subgroups; these results should only be regarded as hypothesis-generating.

Analyses were conducted using Stata version 14.2 (StataCorp).

No formal interim analyses of end points were carried out. At regular intervals, an independent Data Monitoring Committee reviewed confidential reports covering recruitment, safety parameters and primary end point data.

Full details of the statistical methodology are available in the Supplementary Information.

Machine perfusion parameters. During NMP continuous displays of pressures, flows, metabolic (pH) and synthetic (bile production) liver function were available

to the operator. In addition, lactate measurements were carried out using external blood gas analysis. The trial protocol did not stipulate the manner in which these parameters should be interpreted.

Once trial recruitment was complete, an ad hoc analysis was performed in which NMP organs were categorized according to those which, following transplantation, displayed minimal preservation injury (MPI; peak AST < 250 IU l⁻¹) and those with severe preservation injury (SPI; peak AST > 1,000 IU l⁻¹). Groups were compared for differences in donor and recipient characteristics, perfusate biochemistry, bile production and evidence of post-reperfusion syndrome.

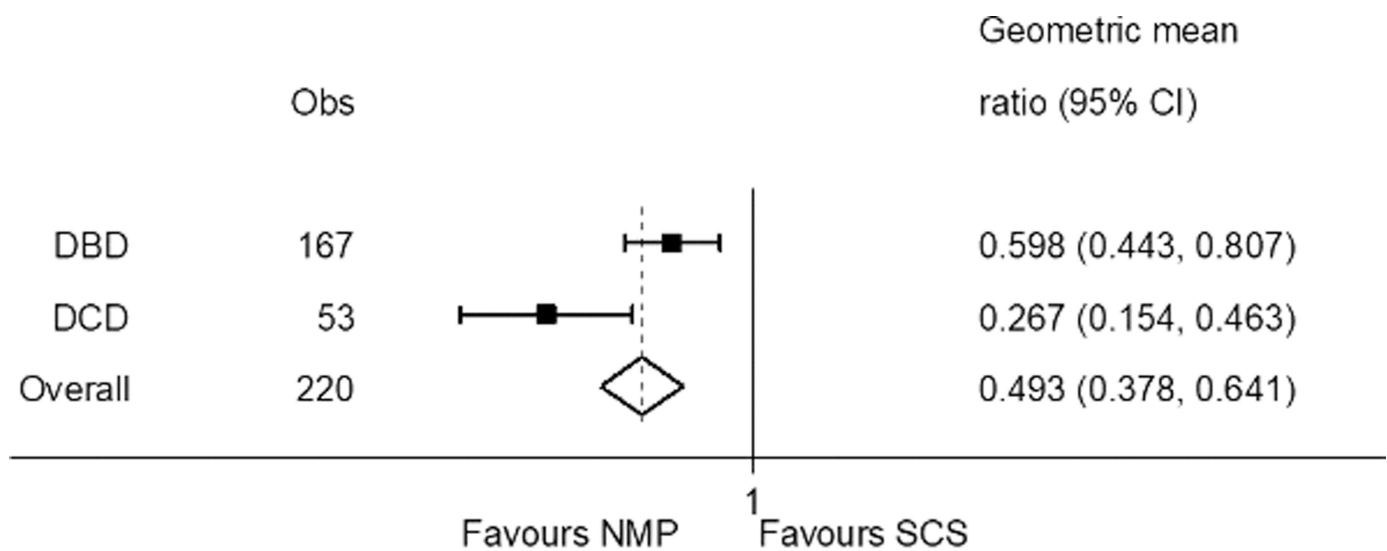
Adverse events. Reporting of adverse events was in accordance with the European Commission MEDDEV guidelines²⁹. Following trial completion, these were reviewed by two independent clinicians blinded to the treatment arm. Adverse events with a Clavien-Dindo³⁰ grading greater than IIIa were considered serious adverse events. Rates of adverse events are reported with 95% confidence intervals. No statistical tests were applied to these data.

Full details of the trial methodology are available in the clinical trial protocol²⁸.

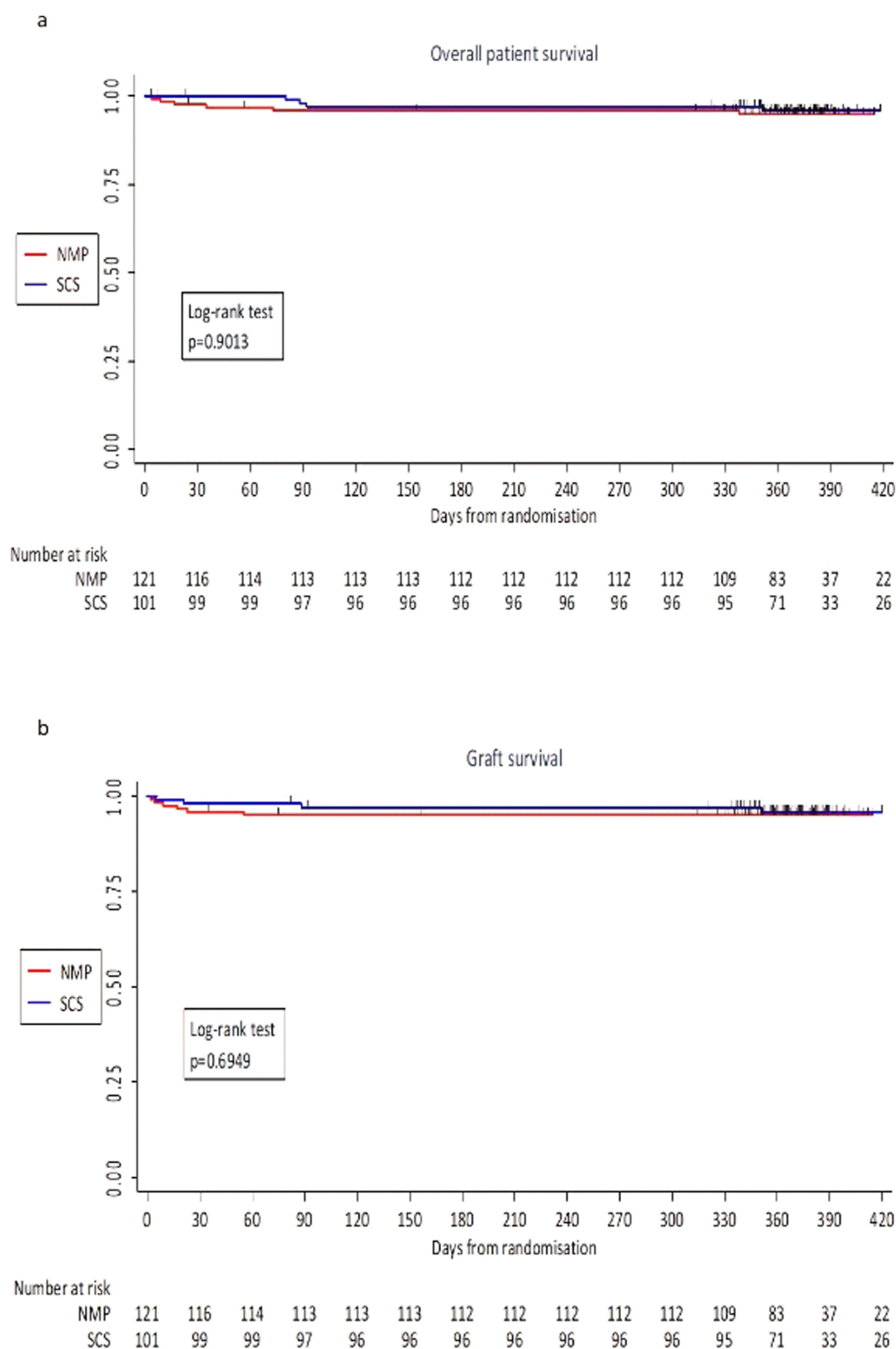
Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request. The full trial protocol, statistical analysis plan and final statistical report are available in the Supplementary Information.

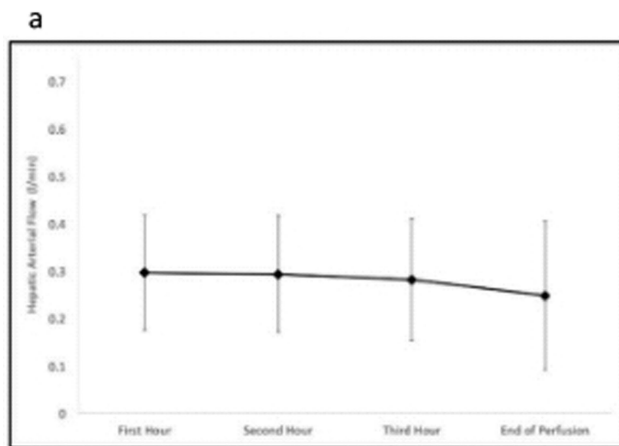
21. Schulz, K. F., Altman, D. G. & Moher, D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br. Med. J* **340**, c332 (2010).
22. Kootstra, G., Daemen, J. H. & Oomen, A. P. Categories of non-heart-beating donors. *Transplant. Proc* **27**, 2893–2894 (1995).
23. Makowka, L. et al. Surgical technique of orthotopic liver transplantation. *Gastroenterol. Clin. North Am* **17**, 33–51 (1988).
24. Glanemann, M. et al. Clinical implications of hepatic preservation injury after adult liver transplantation. *Am. J. Transplant* **3**, 1003–1009 (2003).
25. Gaffey, M. J. et al. Predictive value of intraoperative biopsies and liver function tests for preservation injury in orthotopic liver transplantation. *Hepatology* **25**, 184–189 (1997).
26. Karayalçin, K. et al. The role of dynamic and morphological studies in the assessment of potential liver donors. *Transplantation* **57**, 1323–1327 (1994).
27. Hilmi, I. et al. The impact of postreperfusion syndrome on short-term patient and liver allograft outcome in patients undergoing orthotopic liver transplantation. *Liver Transpl* **14**, 504–508 (2008).
28. Nasralla, D. et al. A multicentre randomised controlled trial to compare the efficacy of ex-vivo normothermic machine perfusion with static cold storage in human liver transplantation. *Protoc. Exch.* <https://doi.org/10.1038/protex.2018.027> (2018).
29. *Guidelines on Medical Devices. Clinical Investigations: Serious Adverse Event Reporting.* Report No. MEDDEV 2.7/3 (European Commission, 2010).
30. Dindo, D., Demartines, N. & Clavien, P. A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann. Surg.* **240**, 205–213 (2004).



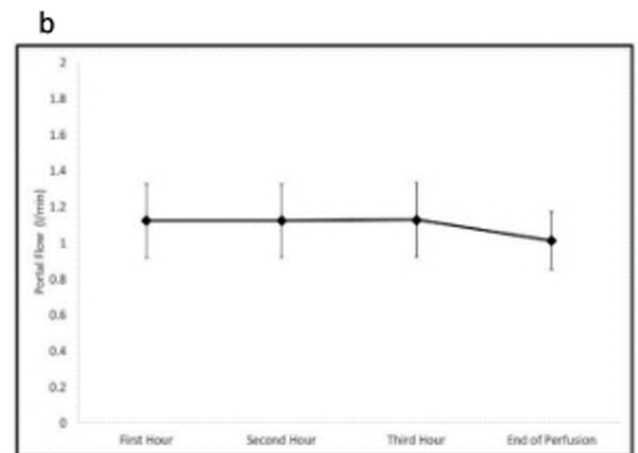
Extended Data Fig. 1 | Forest plot for subgroup analysis of peak AST by donor type. Geometric mean ratio and 95% confidence interval are reported for each subgroup and overall for all groups. DBD group, $n = 87$ NMP, $n = 80$ SCS; DCD group, $n = 33$ NMP, $n = 20$ SCS.



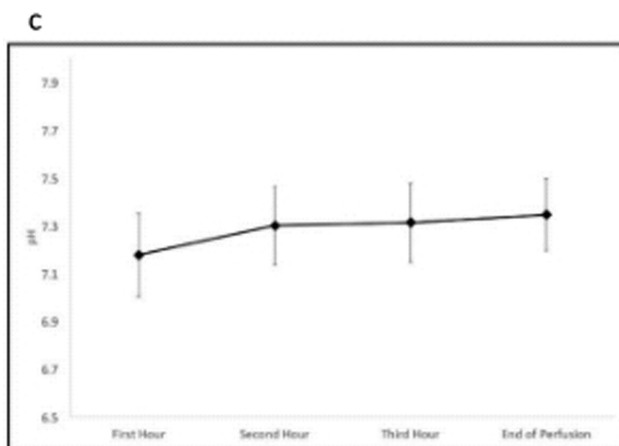
Extended Data Fig. 2 | Post-reperfusion syndrome. **a**, Kaplan–Meier plot for one-year survival of patients with two-sided log-rank test. **b**, Kaplan–Meier plot for one-year graft survival with two-sided log-rank test.



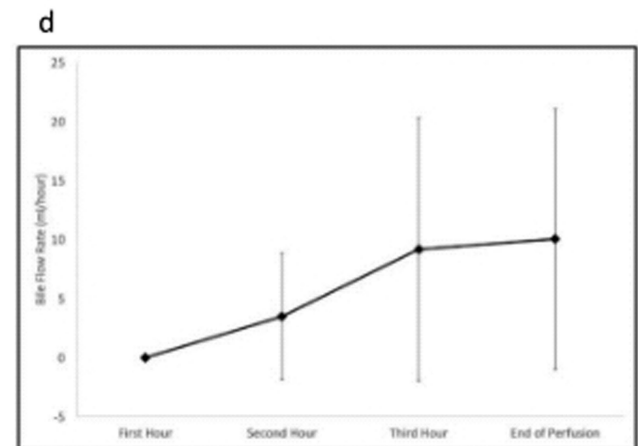
Time Point	Mean flow (litres/min)	Standard Deviation
First hour	0.296	0.121
Second hour	0.294	0.122
Third hour	0.282	0.128
End of perfusion	0.248	0.157



Time Point	Mean flow (litres/min)	Standard Deviation
First hour	1.118	0.205
Second hour	1.118	0.202
Third hour	1.124	0.206
End of perfusion	1.008	0.160



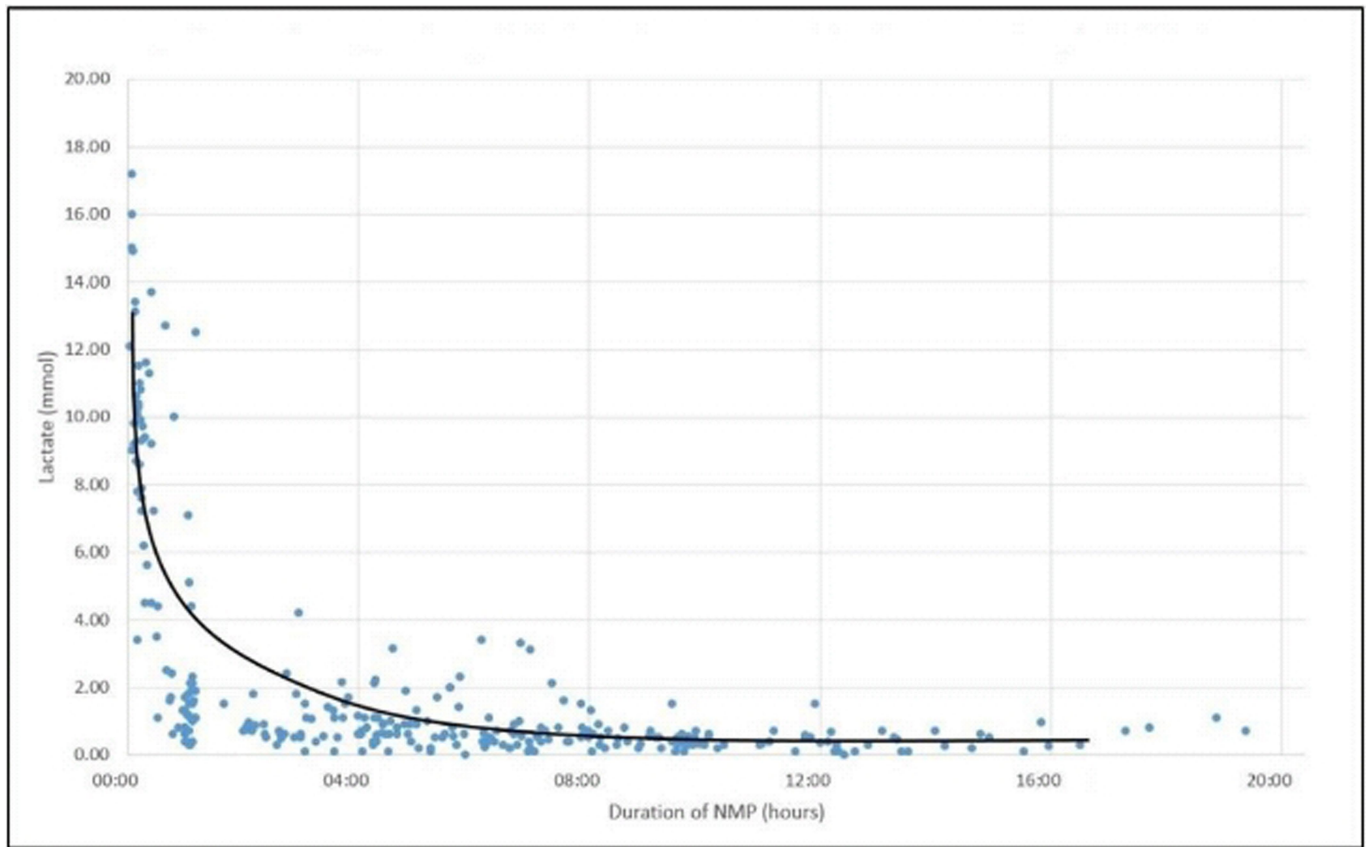
Time Point	Mean pH	Standard Deviation
First hour	7.178	0.174
Second hour	7.301	0.165
Third hour	7.313	0.166
End of perfusion	7.347	0.150



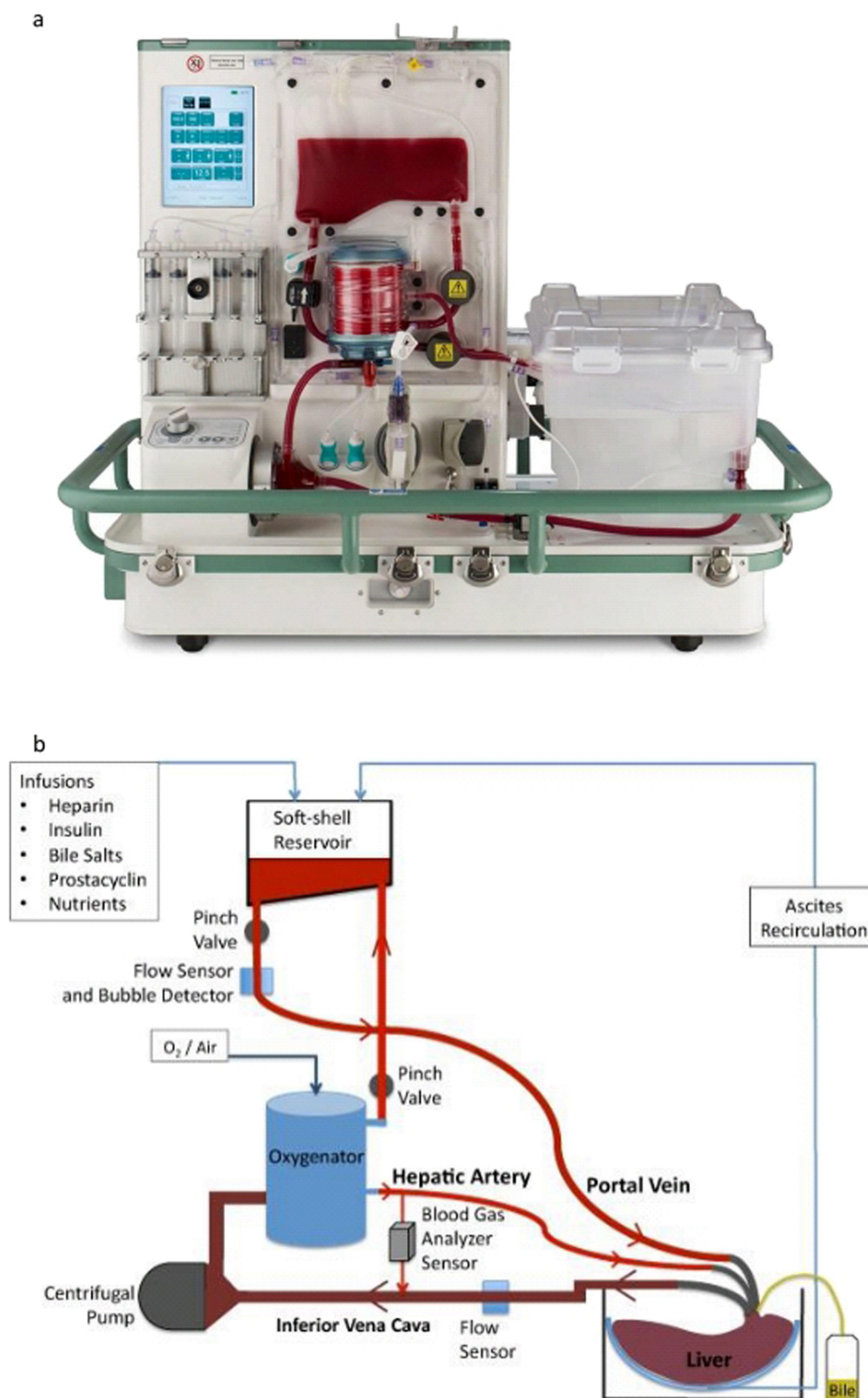
Time Point	Mean flow (ml/hour)	Standard Deviation
First hour	0	0
Second hour	3.486	5.358
Third hour	9.166	11.160
End of perfusion	10.050	11.036

Extended Data Fig. 3 | Machine perfusion parameters during NMP.
a, Hepatic artery flow during NMP. **b**, Portal vein flow during NMP.
c, Perfusate pH during NMP. **d**, Bile production during NMP.

a–d, Data are mean \pm s.d. of each time point. Actual values are shown in the table. $n = 87$.



Extended Data Fig. 4 | Perfusate lactate levels during NMP. Scatter graph with trend line showing perfusate lactate levels at different time points during NMP for all transplanted livers. $n = 94$.



Extended Data Fig. 5 | NMP device and circuit. a, OrganOx metra (generation 1). The NMP device used in the trial. **b,** OrganOx metra NMP circuit. The liver is perfused via the hepatic artery and portal vein. It drains via the inferior vena cava to a centrifugal pump through which the

perfusate passes, via a heat exchanger/oxygenator, to a reservoir or directly into the hepatic artery. The perfusate in the reservoir drains under gravity into the portal vein.

Extended Data Table 1 | Detailed breakdown of reasons for discard of NMP livers

Discarded Liver No	Device Error	Device User Error	Poor Perfusion Parameters	Donor Malignancy	Door Cirrhosis	Poor In-situ Perfusion	Prolonged Donor Warm Ischaemia	Liver Size	Steatosis	Further Details
1	N	Y	N	N	N	N	N	N	Y	Poor perfusion due to IVC cannula positioning. Safely converted to cold storage and discarded due to steatosis.
2	N	N	N	N	Y	N	N	N	Y	Appearances consistent with cirrhosis in donor with known hepatitis C.
3	N	N	Y	N	N	N	N	N	N	Poor hepatic artery flow during NMP with increasing lactate.
4	N	N	N	Y	N	N	N	N	N	Incidental lung tumour found at retrieval
5	N	N	N	N	N	N	Y	N	Y	Warm ischaemia time greater than 30 minutes in DCD donor.
6	N	N	N	N	N	Y	Y	N	Y	Warm ischaemia time greater than 30 minutes in DCD donor. Poor in situ cold perfusion.
7	N	N	Y	N	N	N	N	N	Y	Persistently raised lactate >6mmol after 6 hours NMP.
8	N	N	N	Y	N	N	N	N	Y	Colonic tumour found at retrieval.
9	N	N	N	N	N	N	N	N	Y	60% steatosis on biopsy.
10	N	N	Y	N	N	N	N	N	N	Persistent acidosis with lactate >6mmol after 8hours NMP.
11	N	N	N	N	N	N	N	Y	Y	Large steatotic liver, no size-matched recipient found.
12	N	N	Y	N	N	N	N	N	Y	Persistently raised lactate >3mmol with acidosis.
13	N	N	N	N	N	N	N	N	Y	Moderate steatosis on biopsy, surgeon decision to discard.
14	N	N	Y	N	N	N	N	N	Y	Poor hepatic artery and portal vein flow during NMP in a steatotic liver.
15	N	Y	N	N	N	N	N	N	Y	Excessive bleeding during NMP from phrenic veins and hepatic artery in a steatotic liver. Safely converted to cold storage and declined due to steatosis.
16	Y	N	N	N	N	N	N	N	Y	See Supplementary Information for narrative description of device error.

Extended Data Table 2 | Post-reperfusion syndrome analysis

a

Post-reperfusion syndrome	NMP (n=121)	SCS (n=101)
No	106 (87.6%)	65 (67.0%)
Yes	15 (12.4%)	32 (33.0%)
Total	121	97
Difference = -20.6% (95% C.I. -31.6%, -9.5%) p = 0.000		

b

Post-reperfusion lactate	NMP (n=81)	SCS (n=51)	p-value
Median (IQR)	3.6 (2.6, 4.2)	4.1 (3.2, 5.0)	0.018

c

	NMP (n=121)	SCS (n=101)
Requiring pre-reperfusion vasopressor infusion	92 (76.0%)	82 (81.2%)
<i>missing</i>	2 (1.7%)	6 (5.9%)
Post-reperfusion vasopressor bolus	41 (33.9%)	60 (59.4%)
<i>(missing)</i>	4 (3.3%)	9 (8.9%)
Requiring post-reperfusion vasopressor infusion	65 (53.7%)	80 (79.2%)
<i>(missing)</i>	1 (0.8%)	9 (8.9%)

a. Post-reperfusion syndrome by treatment group. Frequencies and column percentages are reported. Difference in proportions was tested using a Fisher's test for proportions. **b.** Difference in post-reperfusion lactate in the recipient in each treatment arm. This relates to the first lactate measurement recorded by the anaesthetist after liver reperfusion and occurred within 30 min of reperfusion. Analysis using non-parametric Mann–Witney *U*-test. IQR, inter-quartile range. **c.** Difference in use of vasopressor medications before, during and after liver reperfusion in the recipient. Percentage of total events are reported in brackets. Details of the specific vasopressors that were used were not recorded.

Extended Data Table 3 | Extended primary outcome analysis

a

	NMP N=120	SCS N=100	Difference / Mean ratio [^] [% reduction]
Mean In Peak AST (95% C.I.)	6.191 (6.013, 6.368)	6.872 (6.678, 7.066)	-0.681 (-0.946, -0.417)
Geometric Mean Peak AST (95% C.I.)	488.142 (408.856, 582.804)	964.934 (794.471, 1171, 972)	0.506 (0.388, 0.659) [49.4% (34.1%, 61.2%)]

b

Donor Type	Obs	Effect (geometric mean ratio)	[95% Conf. Interval]	p-value
DBD	167	0.598	(0.443, 0.807)	0.001
DCD	53	0.267	(0.154, 0.463)	0.000

a. Primary outcome results from the unadjusted analysis. Sample size for analysis of primary outcome is $n = 220$. A Student's t -test was used. [^]First cell in this column refers to the mean difference in natural logarithm of peak AST (variable used to run the analysis models). The second cell in this column refers to the geometric mean ratio of the peak AST, used to look at the reduction in the original measurement. b. Treatment effect on peak AST for donor type subgroups. Sample size for subgroup analysis is $n = 220$ (DBD group, $n = 87$ NMP, $n = 80$ SCS; DCD group, $n = 33$ NMP, $n = 20$ SCS). A Student's t -test was used. No power calculation or adjustment was made for subgroup analysis.

Extended Data Table 4 | Characteristics and perfusate analysis of livers included in NMP liver quality model development

a

	MPI	SPI	p-value
NMP duration (mins)	592.8 ± 51.10	633.7 ± 52.49	0.579
Donor Type (DBD/DCD)	19/9	20/5	0.365
Donor Age (years)	56.2 ± 3.42	51.0 ± 3.02	0.260
Donor Sex	9M; 19F	18M; 7 F	0.006
ET-DRI	1.83 ± 0.22	1.76 ± 0.16	0.598
Recipient Age	52.2 ± 2.11	56.0 ± 2.77	0.276
Recipient Sex	17M; 11F	19M; 6F	0.257
MELD	15.0 ± 1.09	13.5 ± 1.28	0.392

b

Time (mins)	15 mins			60 mins			End (range 240-1440 mins)			Difference (End – 15mins)		
	MPI	SPI	p-value	MPI	SPI	p-value	MPI	SPI	P value	Change MPI	Change SPI	p-value
Haemolysis Index	0.25	0.19	0.727	0.18	0.21	0.960	0.11	0.38	0.072	-0.04	0.09	0.0278
Urea	3.4 ± 0.19	3.4 ± 0.23	0.994	5.0 ± 0.34	5.0 ± 0.34	0.997	18.9 ± 1.99	18.0 ± 1.43	0.708	15.5 ± 1.93	14.5 ± 1.37	0.6880
Bilirubin	2 (2)	2 (2-10)	0.034	2 (2-5)	2 (2-28)	0.026	2 (2-108)	4.5 (2-121)	0.102	0	2.5	0.1443
ALT	170.5 (64-1811)	669 (58-4390)	0.005	193.5 (78-2306)	570 (78-4809)	0.006	268.5 (106-4107)	1334 (266-16772)	<0.001	56	461	0.0000
Alkaline Phosphatase	7 (5-21)	7 (5-32)	>0.999	7 (5-45)	10 (5-105)	0.113	40 (5-394)	71.5 (8-322)	0.051	32	61.5	0.0682
GGT	4 (4-30)	4.5 (4-92)	0.262	5 (4-103)	12 (4-110)	0.132	35.5 (4-190)	108 (8-759)	0.008	23 (0-183)	104 (4-667)	0.0039
LDH	1073 ± 180	1838 ± 245	0.015	1491 ± 256.6	1884 ± 207.8	0.257	1479 ± 157.2	2610 ± 187.6	<0.001	482.8 ± 110.6	980.1 ± 278.7	0.0591
CRP	8.3 ± 0.87	8.6 ± 1.47	0.857	22.1 ± 2.83	23.0 ± 3.63	0.853	147.4 ± 24.39	191.5 ± 36.07	0.300	131.9 ± 27.28	189.2 ± 42.74	0.2445
Lactate	10.5 ± 0.61	10.7 ± 0.78	0.852	1.7 ± 0.39	3.0 ± 0.78	0.106	0.5 ± 0.09	0.9 ± 0.29	0.146	-8.6 ± 0.64	-7.7 ± 0.83	0.3672

a. Demographic data for organ quality model development livers. Demographic data for minimal preservation injury (MPI; peak AST < 250 IU l⁻¹; n = 28) and significant preservation injury (SPI; peak AST > 1,000 IU l⁻¹; n = 25) groups. Continuous variables were analysed using an unpaired Student's t-test and categorical variables using Fisher's exact test. Data are mean ± s.e.m. b. Comparison of NMP perfusate analyses between MPI (peak AST < 250 IU l⁻¹; n = 28) and SPI (peak AST > 1,000 IU l⁻¹; n = 25) groups. A D'Agostino-Pearson normality test was performed to assess data distribution. Parametric data were analysed using an unpaired Student's t-test and non-parametric data were analysed using a Mann-Whitney U-test. Parametric data are presented as mean ± s.e.m. and non-parametric data are presented as median and range.

Extended Data Table 5 | Adverse events analysis

a

Patients with	NMP (N=121)	SCS (N=101)	Total
No events reported	54 (44.6%)	43 (42.6%)	97 (43.7%)
Adverse events (95% C.I.)	67 (55.4%) (46.1%, 64.4%)	58 (57.4%) (47.2%, 67.2%)	125 (56.3%)

b

Clavien-Dindo grading	NMP	SCS	Total
I	15 (11.7%)	30 (18.3%)	45 (15.4%)
II	64 (50.0%)	72 (43.9%)	136 (46.6%)
IIIa	28 (21.9%)	26 (15.9%)	54 (18.5%)
IIIb	8 (6.3%)	9 (5.5%)	17 (5.8%)
IVa	5 (3.9%)	15 (9.2%)	20 (6.9%)
IVb	3 (2.3%)	9 (5.5%)	12 (4.1%)
V	5 (3.9%)	3 (1.8%)	8 (2.7%)
Total	128	164	292

c

Classification	NMP	SCS	Total
AE	107 (83.6%)	128 (78.1%)	235 (80.5%)
SAE	21 (16.4%)	36 (22.0%)	57 (19.5%)
Total	128	164	292

a, Number of patients with any adverse events reported in each trial arm. The percentage of total events is reported in brackets. No statistical tests have been applied. **b**, Adverse events were categorized by Clavien-Dindo grade. Breakdown of adverse events in each trial arm according to Clavien-Dindo grading. The percentage of total events is reported in brackets. Adverse events with Clavien-Dindo grading \geq IIIb were categorized as serious adverse events. No statistical tests have been applied. **c**, Breakdown of adverse events and serious adverse events in each trial arm. The percentage of total events is reported in brackets. Adverse events with Clavien-Dindo grading \geq IIIb were categorized as serious adverse events. No statistical tests have been applied.

Extended Data Table 6 | Detailed breakdown of adverse events in each trial arm

Event Category	NMP	SCS	Total
Infection	25 (19.5%)	17 (10.4%)	42 (14.4%)
Chest	1	1	2
Blood	10	3	13
Biliary	6	0	6
Abdominal	2	3	5
Gastrointestinal	4	5	9
Other	2	5	7
Hepatic	44 (34.4%)	48 (29.3%)	92 (31.5%)
Bile leak	2	1	3
Biliary stricture (anastomotic)	9	11	20
Ischaemic cholangiopathy	1	3	4
Biliary other	1	0	1
Drainage of ascites	0	1	1
Hepatic artery aneurysm	0	1	1
Hepatic artery thrombosis	2	4	6
Hepatic artery stenosis	5	3	8
Hepatic artery other	0	2	2
Hepatic vein thrombosis	1	0	1
Portal vein thrombosis	2	0	2
Portal vein stenosis	2	0	2
Portal vein other	1	0	1
Graft dysfunction	3	2	5
Rejection	12	13	25
Other	3	7	10
Cardiovascular	5 (3.9%)	5 (3.1%)	10 (3.4%)
Congestive heart failure	1	0	1
Myocardial infarction	2	3	5
Other	2	2	4
Dermatologic	1 (0.8%)	0 (0.0%)	1 (0.3%)
Seroma	1	0	1
Gastrointestinal	5 (3.9%)	6 (3.7%)	11 (3.8%)
Colitis	0	1	1
Diarrhea	3	2	5
Other	2	3	5
Genitourinary	8 (6.3%)	17 (10.4%)	25 (8.6%)
Renal insufficiency	6	13	19
UTI	2	3	5
Other	0	1	1
Respiratory	4 (3.1%)	9 (5.5%)	13 (4.5%)
Cold/flu	0	1	1
Pneumonia	4	6	10
Shortness of breath	0	1	1
Other	0	1	1
Bleeding complications	9 (7.0%)	6 (3.7%)	15 (5.1%)
Bleeding – no transfusion required	0	2	2
Hemorrhage (Bleeding requiring transfusion)	3	0	3
Bleeding from hepatic artery	1	1	2
Bleeding from liver parenchyma	2	0	2
Other	3	3	6
Fluid Collection	7 (5.5%)	18 (11.0%)	25 (8.6%)
Abdominal	5	10	15
Pleural	2	7	9
Other	0	1	1
Device error	1 (0.8%)	-	1 (0.3%)
Device user error	2 (1.6%)	-	2 (0.7%)
Other systemic diseases	17 (13.3%)	38 (23.2%)	55 (18.8%)
Total	128	164	292

The percentage of total events is reported in brackets. No statistical tests have been applied.

SAMHD1 acts at stalled replication forks to prevent interferon induction

Flavie Coquel^{1,12}, Maria-Joao Silva^{1,11,12}, Hervé Técher^{2,12}, Karina Zadorozhny³, Sushma Sharma⁴, Jadwiga Nieminiuszcz⁵, Clément Mettling⁶, Elodie Dardillac⁷, Antoine Barthe¹, Anne-Lyne Schmitz¹, Alexy Promonet¹, Alexandra Cribier⁸, Amélie Sarrazin⁹, Wojciech Niedzwiedz⁵, Bernard Lopez⁷, Vincenzo Costanzo², Lumir Krejci^{3,10}, Andrei Chabes⁴, Monsef Benkirane⁸, Yea-Lih Lin^{1,13*} & Philippe Pasero^{1,13*}

SAMHD1 was previously characterized as a dNTPase that protects cells from viral infections. Mutations in SAMHD1 are implicated in cancer development and in a severe congenital inflammatory disease known as Aicardi–Goutières syndrome. The mechanism by which SAMHD1 protects against cancer and chronic inflammation is unknown. Here we show that SAMHD1 promotes degradation of nascent DNA at stalled replication forks in human cell lines by stimulating the exonuclease activity of MRE11. This function activates the ATR–CHK1 checkpoint and allows the forks to restart replication. In SAMHD1-depleted cells, single-stranded DNA fragments are released from stalled forks and accumulate in the cytosol, where they activate the cGAS–STING pathway to induce expression of pro-inflammatory type I interferons. SAMHD1 is thus an important player in the replication stress response, which prevents chronic inflammation by limiting the release of single-stranded DNA from stalled replication forks.

SAMHD1 (sterile alpha motif and HD domain-containing protein 1) is a dNTPase that, in quiescent cells, restricts infection by HIV-1 and other viruses^{1–3}. In cycling cells, this dNTPase activity is inhibited by the phosphorylation of SAMHD1 by cyclin-dependent kinases^{4,5}. Germline mutations in *SAMHD1* cause Aicardi–Goutières syndrome (AGS), a rare inflammatory encephalopathy characterized by overproduction of type I interferons (IFNs)⁶. Other genes mutated in AGS include *TREX1*, which encodes a 3′–5′ exonuclease that degrades nucleic acids in the cytoplasm^{6,7}. Cytosolic DNA species that are not degraded by TREX1 trigger the production of type I IFNs and other cytokines through the cGAS–STING cytosolic DNA-sensing pathway⁸. The mechanism by which SAMHD1 inhibits this pathway is currently unclear^{9–11}, but it has been proposed to involve an elusive 3′–5′ exonuclease activity^{12–14}.

Besides AGS, *SAMHD1* is frequently mutated in solid tumours and in chronic lymphocytic leukaemia^{15,16}. Because SAMHD1 regulates intracellular dNTP pools, imbalanced dNTP levels in SAMHD1-deficient cells might perturb the progression of replication forks and thus increase spontaneous mutagenesis^{16,17}. Moreover, SAMHD1 colocalizes with DNA repair foci in cells exposed to genotoxic agents¹⁵, suggesting that it may have a more direct role at DNA lesions or at stalled replication forks. Fork stalling occurs when cells are exposed to genotoxic agents such as hydroxyurea (HU) and camptothecin (CPT), or when they encounter sequences that are intrinsically difficult to replicate¹⁸. Single-stranded DNA (ssDNA) accumulates at stalled forks and recruits the checkpoint kinase ATR, which in turn activates CHK1 to promote fork restart and prevent premature entry into mitosis^{18,19}. Fork restart depends on the degradation of nascent DNA strands by MRE11 through a process regulated by

BRCA2 and known as fork resection^{20–22}. Defective fork processing leads to fork collapse, increased genomic instability and cancer development^{18,19}.

SAMHD1 prevents release of ssDNA

To determine whether SAMHD1 has a role at stalled forks that could be important to prevent the accumulation of cytosolic DNA, we first constructed stable HEK293T cell lines expressing either a short hairpin RNA (shRNA) against *SAMHD1* (shSAM) or a scrambled shRNA control (shScr; Extended Data Fig. 1a). Confocal immunofluorescence microscopy confirmed the presence of cytosolic ssDNA in shSAM cells, but at a much lower level than in *TREX1*-depleted HEK293T cells, used here as positive control (Fig. 1a, b and Extended Data Fig. 1b). SAMHD1 depletion also induced type I IFN and other pro-inflammatory cytokine mRNAs in HEK293T (Fig. 1c), HL116 cells (Extended Data Fig. 1c) and HeLa cells (Extended Data Fig. 1d). Importantly, both the accumulation of cytosolic ssDNA and the induction of IFN genes were substantially increased when shSAM cells were exposed to HU to induce replication fork stalling (Fig. 1a, b, d and Extended Data Fig. 1e). Moreover, when newly replicated DNA was labelled with BrdU two hours before the addition of HU, this cytosolic DNA contained BrdU (Fig. 1e and Extended Data Fig. 1f), indicating that it comes from arrested forks and not, for example, from damaged mitochondria—another possible source of cytosolic DNA (Extended Data Fig. 1g, h). This induction of IFN genes by SAMHD1 depletion in HeLa, HEK293 and THP1 cells^{23,24} is mediated by the cGAS–STING pathway (Fig. 1f and Extended Data Fig. 1i, j, n, o), as is the case in *Samhd1*^{−/−} mice⁹. It also depends on the interferon regulatory factor IRF3 (Extended Data Fig. 1k–n), both in the presence or the absence of HU. Together, these

¹Institut de Génétique Humaine, CNRS, Université de Montpellier, Laboratoire Maintien de l'Intégrité du Génome au cours de la Réplication, Ligue Contre le Cancer, Montpellier, France. ²IFOM, The FIRC Institute of Molecular Oncology, Milan, Italy. ³Department of Biology and National Centre for Biomolecular Research, Masaryk University, Brno, Czech Republic. ⁴Department of Medical Biochemistry and Biophysics and Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden. ⁵The Institute of Cancer Research, London, UK. ⁶Institut de Génétique Humaine, CNRS, Université de Montpellier, Domiciliation, Activation Immunitaire et Infection, Montpellier, France. ⁷Université Paris Sud, CNRS, UMR 8200 and Institut de Cancérologie Gustave Roussy, Ligue Contre le Cancer, Villejuif, France. ⁸Institut de Génétique Humaine, CNRS, Université de Montpellier, Laboratoire de Virologie Moléculaire, Montpellier, France. ⁹BioCampus Montpellier, Université de Montpellier, CNRS, Montpellier, France. ¹⁰International Clinical Research Center, St Anne's University Hospital, Brno, Czech Republic. ¹¹Present address: Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. ¹²These authors contributed equally: Flavie Coquel, Maria-Joao Silva, Hervé Técher. ¹³These authors jointly supervised this work: Yea-Lih Lin, Philippe Pasero. *e-mail: ylin@igh.cnrs.fr; ppasero@igh.cnrs.fr

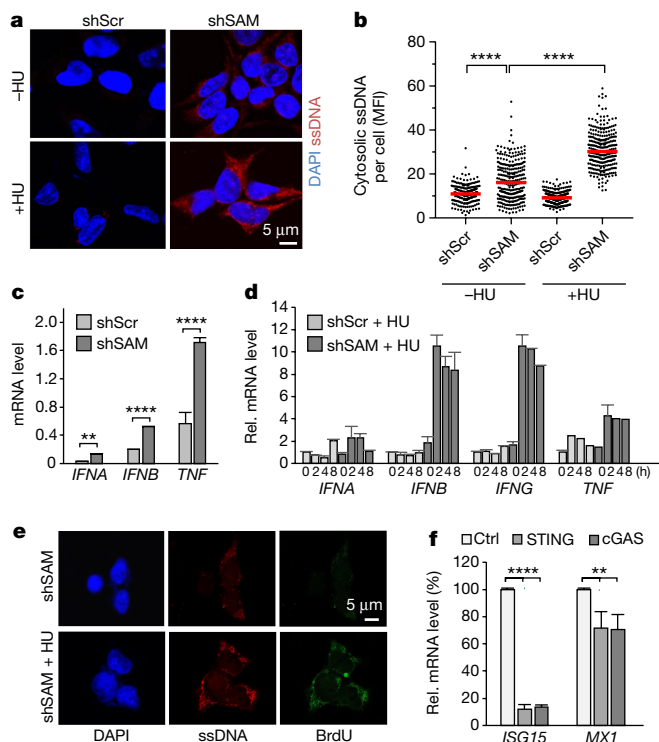


Fig. 1 | SAMHD1 prevents accumulation of cytosolic ssDNA and induction of type I IFNs in response to replication stress. **a**, Cytosolic DNA (red) in control (shScr) and SAMHD1-depleted (shSAM) HEK293T cells exposed for 2 h to 4 mM HU. **b**, Mean fluorescence intensity (MFI) of cytosolic ssDNA per cell ($n = 300$). **c**, Expression of *IFN* and *TNF* mRNAs in shScr and shSAM HEK293T cells (mean and s.d. of three independent experiments). **d**, Induction of *IFN* and *TNF* mRNAs in shScr and shSAM HEK293T cells exposed to HU as in **a**. Mean and s.d. correspond to technical triplicates, representative of three independent experiments. **e**, Cytosolic DNA and BrdU-labelled DNA in shSAM HEK293T cells incubated with BrdU for 2 h then without BrdU and in the presence of 4 mM HU for 2 h ($n = 3$). **f**, Expression of IFN-induced genes *ISG15* and *MX1* in cGAS-knockout (cGAS), STING-knockout (STING) and control (ctrl) SAMHD1-depleted HeLa cells. Mean and s.d. of three independent experiments are shown. **** $P < 0.0001$, ** $P < 0.01$, two-sided Mann–Whitney test.

data indicate that SAMHD1 prevents the release of ssDNA from stalled forks and aberrant activation of the cGAS–STING pathway.

SAMHD1 is involved in DNA replication

To investigate the role of SAMHD1 in DNA replication, we first monitored its subcellular localization by immunofluorescence microscopy. SAMHD1 forms foci in HeLa cell nuclei that colocalize with replication sites (Fig. 2a and Extended Data Fig. 2a). By using the iPOND (isolation of proteins on nascent DNA) method²⁵, we confirmed that SAMHD1 is present at replication forks (Fig. 2b and Extended Data Fig. 2b). However, SAMHD1 persisted on newly replicated chromatin after a thymidine chase, unlike PCNA (Fig. 2b). Samhd1 also interacted with nascent DNA in *Xenopus* egg extracts and was recruited to chromatin in response to replication inhibition and DNA double-strand breaks (DSBs; Extended Data Fig. 2d–f).

SAMHD1-depleted cells grew more slowly than control cells (Extended Data Fig. 3a, b) and had a longer S phase (Extended Data Fig. 3c). To determine whether this was due to slower DNA synthesis, we labelled control and shSAM cells with IdU and CldU and monitored the progression of individual forks by DNA fibre spreading. CldU tracks were much shorter in HEK293T and HeLa cells depleted of SAMHD1 than they were in control cells (Fig. 2c and Extended Data Fig. 3d–f), indicating slowed DNA synthesis. Moreover, increased fork stalling was detected in shSAM cells by monitoring the asymmetry of

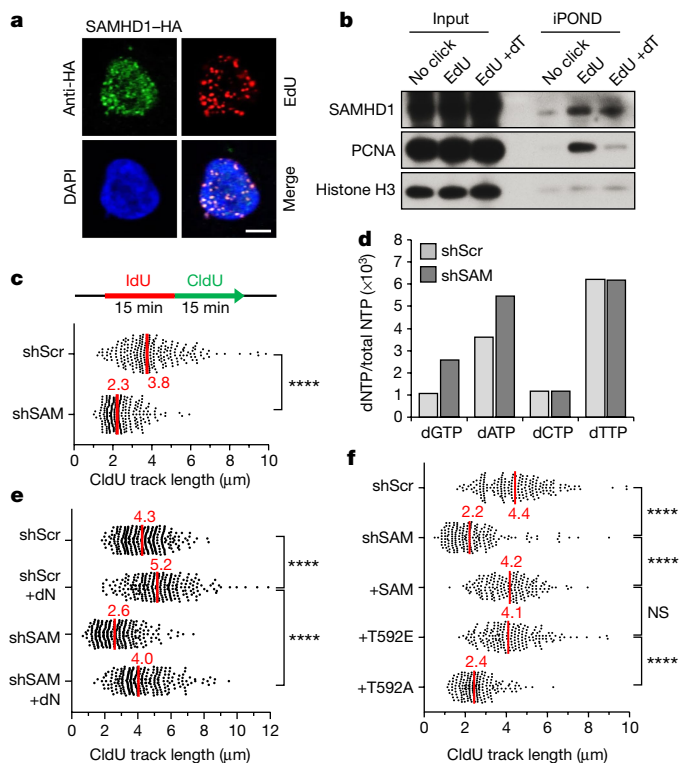


Fig. 2 | SAMHD1 colocalizes with replication foci and regulates fork progression. **a**, Immunolocalization of haemagglutinin (HA)-tagged SAMHD1 (green) and EdU-labelled replication foci (red) in HeLa cells, showing their colocalization (merge) in a DAPI-stained nucleus (blue; $n = 2$). **b**, iPOND analysis of SAMHD1 and PCNA in HEK293T cells. **c**, DNA fibre analysis of CldU track lengths ($n = 250$) in shScr and shSAM HEK293T cells. **d**, Intracellular dNTP pools in shScr and shSAM HEK293T cells ($n = 2$). **e**, Lengths of CldU tracks ($n = 250$) after addition of a balanced mix of nucleosides (+dN) 2 h before IdU and CldU labelling. **f**, Fork progression in shSAM HEK293T cells complemented with wild-type SAMHD1 or with phosphomimetic (+T592E) or non-phosphorylatable (+T592A) mutants ($n = 200$). In **c**, **e** and **f**, median track lengths are indicated in red. **** $P < 0.0001$, two-sided Mann–Whitney test. NS, not significant.

sister forks (Extended Data Fig. 3g), even though the density of active forks was unchanged ($P = 0.33$, Extended Data Fig. 3h). Together, these data indicate that SAMHD1 promotes normal fork progression.

To determine whether the replication defects observed in shSAM cells were due to imbalanced dNTP pools, we measured dNTP levels in shSAM HEK293T cells by HPLC (Fig. 2d and Extended Data Fig. 2c). In SAMHD1-depleted cells, dGTP levels were 2.5-fold higher than in controls, but dATP levels were only approximately 30% higher and dCTP and dTTP levels were unchanged. Moreover, the addition of a balanced mix of nucleosides to shSAM cells only partially rescued fork speed (Fig. 2e), suggesting that the role of SAMHD1 in S phase is not limited to the regulation of dNTP pools.

SAMHD1 is phosphorylated on T592 by cyclin-dependent kinases (CDKs) during S and G₂/M phases of the cell cycle^{4,5} (Extended Data Fig. 4a). To determine how this phosphorylation affects DNA replication, shSAM cells were complemented with phosphomimetic (T592E) or non-phosphorylatable (T592A) SAMHD1 mutants (Extended Data Fig. 4a). Unlike the T592A mutant, the T592E mutant fully rescued slow forks in shSAM cells (Fig. 2f). Because both cell types have similar dNTP levels (Extended Data Fig. 4b), these data indicate that phosphorylation of SAMHD1 on T592 promotes fork progression independently of dNTP pools. Interestingly, normal fork progression was also restored to shSAM cells by the expression of the dNTPase-deficient K312A mutant but not by the dNTPase-proficient Y315A mutant¹³ (Extended Data Fig. 4a, c). Fork progression was also impaired in

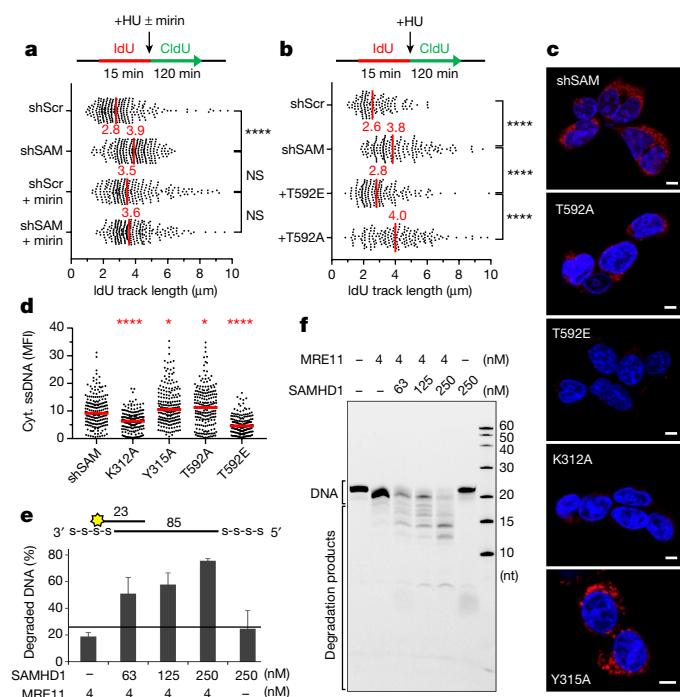


Fig. 3 | SAMHD1 promotes the MRE11-dependent degradation of nascent DNA at stalled forks. **a**, shScr and shSAM HEK293T cells were labelled with IdU and CldU in the presence of 4 mM HU and in the presence or absence of 50 μM mirin. Median IdU track lengths ($n = 240$) are indicated in red. **** $P < 0.0001$, Mann–Whitney test. **b**, Analysis as in **a** of IdU track lengths ($n = 180$) in shSAM HEK293T cells expressing the T592A and T592E mutants of SAMHD1. **c**, Cytosolic ssDNA (red) in shSAM HEK293T cells ($n = 200$) complemented with K312A, Y315A, T592A and T592E mutants and exposed for 2 h to 4 mM HU. Scale bars, 5 μm. **d**, Mean fluorescence intensity of cytosolic ssDNA per cell ($n = 200$). Relative difference to shSAM: * $P < 0.05$, **** $P < 0.0001$, Mann–Whitney test. **e**, In vitro nuclease assays of MRE11 in the absence and presence of increasing concentrations of SAMHD1. Mean and s.d. are from three independent experiments. **f**, Representative gel as in **e**.

immortalized B cells from a patient with AGS (*SAMHD1*^{−/−}) with a Q548X mutation (Extended Data Fig. 4d), which does not affect dNTP regulation^{12,14}. Together, these data indicate that phosphorylation of SAMHD1 on T592 by CDK is required for normal fork progression.

SAMHD1 promotes fork resection

Because SAMHD1 was proposed to have potential 3′–5′ exonuclease activity^{13,14} similar to MRE11, we next asked whether it could be involved in fork resection. We labelled newly synthesized DNA in control and shSAM cells with IdU for 15 min then exposed them to HU for 120 min in the presence of CldU. In control cells, a significant degradation of IdU tracks was observed after HU addition (Extended Data Fig. 5a), indicating resection. This was prevented by SAMHD1 depletion and/or by addition of the MRE11 inhibitor mirin^{20,21} (Fig. 3a). Similar results were obtained by using a different labelling strategy (Extended Data Fig. 5b) and aphidicolin (a drug that inhibits DNA polymerases without affecting dNTP pool) instead of HU to stall replication forks (Extended Data Fig. 5c). Moreover, the addition of exogenous nucleosides did not restore fork resection in shSAM cells (Extended Data Fig. 5d) and dNTP pools were equally affected by HU exposure in control and shSAM cells (Extended Data Fig. 5e). Fork resection was also inhibited by the depletion of CtIP, a cofactor of MRE11 (Extended Data Fig. 5f) and by depleting SMARCA1 (Extended Data Fig. 5g). This suggests that resection occurs partly on reversed forks, which form upon reannealing of nascent DNA strands^{26–28}.

To investigate whether SAMHD1 is also required to resect DNA ends during DSB repair, we used the single-molecule analysis of resection

tracks (SMART) method²⁹ and an in vivo DNA repair assay³⁰ in shSAM and control cells. These approaches demonstrated that SAMHD1 is indeed required to resect DNA ends and repair DSBs (Extended Data Fig. 6a–e), confirming a recent report³¹. Also, we showed that *Xenopus* Samhd1 is recruited to chromatin after EcoRI digestion (Extended Data Fig. 6f). Altogether, these data indicate that SAMHD1 acts with MRE11 and CtIP to promote resection at forks and DSBs, independently of dNTP pools.

Fork resection in shSAM cells was fully restored by the phosphomimetic T592E mutant of SAMHD1, but not by the non-phosphorylatable T592A mutant (Fig. 3b) and it was restored by the dNTPase-defective mutant K312A, but not by the dNTPase-proficient mutant Y315A (Extended Data Fig. 5h), suggesting the resection function of SAMHD1 requires phosphorylation but not the dNTPase activity. Also, fork resection in immortalized B cells was impaired by the Q548X mutation found in a patient with AGS (Extended Data Fig. 5i). The resection-proficient mutants, T592E and K312A, prevented the accumulation of cytosolic DNA whereas the resection-deficient mutants, T592A and Y315A, did not (Fig. 3c, d). Altogether, these data indicate that mutations in SAMHD1 affecting fork progression and nascent DNA resection are also associated with the release of ssDNA fragments into the cytoplasm.

SAMHD1 activates the MRE11 exonuclease

To characterize further the function of SAMHD1 at the biochemical level, we assayed the binding of human SAMHD1 to various DNA substrates (Extended Data Fig. 7a, b) and found a high affinity for ssDNA and different fork structures. Remarkably, SAMHD1 stimulated three-fold the exonuclease activity of MRE11 (Fig. 3e, f), whereas it did not increase the activity of DNA2, FEN1 and bacterial ExoIII (Extended Data Fig. 7c). We assayed direct binding of SAMHD1 to labelled MRE11 by using microscale thermophoresis and measured a dissociation constant (K_d) of 977 ± 176 nM (mean \pm s.e.m.), suggesting that stimulation of the exonuclease activity of MRE11 by SAMHD1 is due to a direct interaction (Extended Data Fig. 7d). In addition to binding to MRE11, we found that SAMHD1 directly interacts with replication protein A (RPA) with a $K_d = 312 \pm 57$ nM as determined by microscale thermophoresis (Extended Data Fig. 7e) and by co-immunoprecipitation in *Xenopus* egg extracts (Extended Data Fig. 7f). *Xenopus* Samhd1 also interacted with the resection protein CtIP (Extended Data Fig. 7g). Together, these data suggest that SAMHD1 binds with high affinity to RPA and fork structures, where it interacts directly with MRE11 to selectively stimulate its exonuclease activity.

SAMHD1 activates CHK1 at stalled forks

Because SAMHD1 binds to forks, MRE11 and RPA in vitro, and MRE11 is reported to have a role in CHK1 activation at stalled forks³², we asked whether the recruitment of MRE11 and the formation of RPA-coated ssDNA at stalled forks depends on SAMHD1. To address this possibility, we first examined the colocalization of MRE11 and EdU-labelled replication sites in HU-treated shSAM and control cells and found that their colocalization requires SAMHD1 (Extended Data Fig. 8a). To test whether the formation of RPA-coated ssDNA also depends on SAMHD1, we looked for ssDNA in HU-treated shSAM and control cells by detecting BrdU incorporation under non-denaturing conditions. Both ssDNA formation (Extended Data Fig. 8b) and the formation of RPA foci (Extended Data Fig. 8c) depended on the presence of SAMHD1. Incidentally, we also observed a fourfold increase in BrdU-labelled cytosolic ssDNA in shSAM cells (Extended Data Fig. 8b, asterisks), consistent with the data shown in Fig. 1a, b.

Because RPA recruits ATR to stalled forks, we next monitored the phosphorylation of the ATR targets CHK1 and H2AX in cells exposed either to HU or to CPT (Extended Data Fig. 8d, e). Phosphorylation of both proteins was much reduced in shSAM cells when compared to control cells, indicating that activation of the ATR–CHK1 pathway depends on SAMHD1. CHK1 phosphorylation was restored upon complementation of shSAM cells with wild-type SAMHD1 (Extended Data

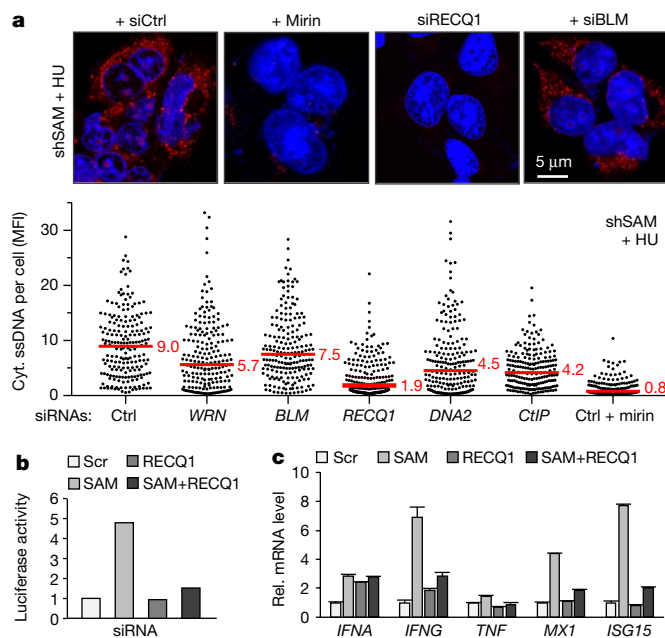


Fig. 4 | Depletion of RECQ1 prevents the IFN response in SAMHD1-depleted cells. **a**, Cytosolic ssDNA in HU-treated shSAM HEK293T cells transfected for 48 h with short interfering RNAs (siRNAs) against WRN, BLM, RECQ1 (also known as RECQL), DNA2 and CtIP (also known as RBBP8) or treated with mirin ($n = 200$; two experiments). **b**, Induction of IFNs (luciferase assay) in HL116 cells transfected with siRNAs against SAMHD1 (SAM), RECQ1 (RECQ1) or both (SAM + RECQ1), or with a control scrambled siRNA (Scr; $n = 3$). Data are representative of three experiments. **c**, Total mRNA was extracted from HL116 cells after siRNA transfection and the expression of *IFNA*, *IFNG*, *TNF*, *MX1* and *ISG15* was quantified by qRT-PCR ($n = 3$). Error bars denote s.d. of triplicates from representative experiments.

Fig. 8f) or with the T592E and K312A mutants, but not with T592A and Y315A mutants, which are defective in fork resection (Extended Data Fig. 8g, h). Moreover, the activation of Chk1 in *Xenopus* egg extracts treated with aphidicolin or DSB-inducing agents was also inhibited in the absence of Samhd1 or Mre11 activity (Extended Data Fig. 8i–k). We also found that SAMHD1 acts together with MRE11 for the restart of CPT-arrested forks (Extended Data Fig. 8l). Together, these findings indicate a novel role for SAMHD1 in the activation of the ATR–CHK1 pathway and in fork restart.

RECQ1 induces IFNs in shSAM cells

We hypothesized that the cytosolic ssDNA observed in shSAM cells might result from the displacement of nascent DNA by a helicase and cleavage by a flap endonuclease. To test this hypothesis, we depleted shSAM cells of BLM, WRN, RECQ1, DNA2 and CtIP (Extended Data Fig. 9a) and exposed them to HU. Depletion of RECQ1 and, to a lesser extent, DNA2 and CtIP resulted in a significant reduction in the cytosolic ssDNA seen in both HU-treated and untreated shSAM cells (Fig. 4a and Extended Data Fig. 9b–e). RECQ1 depletion also reduced the expression of IFN genes in HL116 (Fig. 4b, c), HEK293T and HeLa cells (Extended Data Fig. 9f, g) and the degradation of IdU tracks (Extended Data Fig. 9h), suggesting that RECQ1 contributes to fork resection and is responsible for the observed cytosolic ssDNA. Levels of cytosolic ssDNA were also reduced by mirin (Fig. 4a), suggesting that the endonuclease activity of MRE11 cleaves these displaced strands, releasing short ssDNA fragments, as it does at DSBs³³.

To investigate further how cytosolic ssDNA is produced, shSAM cells were labelled with IdU before the addition of HU and with CldU during the HU arrest, and then released from HU in the presence of thymidine to chase CldU. Remarkably, both IdU- and CldU-labelled ssDNA were detected in the cytoplasm (Extended Data Fig. 10a, b),

indicating that DNA synthesized both before and after replication fork stalling contributes to the cytosolic ssDNA seen in HU-treated cells (Extended Data Fig. 10c).

We conclude from our data that SAMHD1 functions during S phase at stalled DNA replication forks to activate the ATR–CHK1 pathway and to promote the resection of gapped or reversed forks by binding to and activating the MRE11 exonuclease. This function is independent of its previously described role in maintaining balanced dNTP pools, which is also important for accurate DNA replication^{16,17}. This novel role of SAMHD1 depends on its phosphorylation by S-phase CDKs, which act as a switch to control its dNTPase-dependent and -independent functions.

In the absence of SAMHD1, ssDNA molecules are released from stalled replication forks and accumulate in the cytoplasm, where they activate the cGAS–STING pathway, inducing expression of type I IFNs and other pro-inflammatory cytokine genes (Extended Data Fig. 10d). This is consistent with recent reports showing that DNA repair byproducts can also induce the production of type I IFNs^{34–37}. Mutations in SAMHD1 that prevent the degradation of nascent DNA at stalled forks increase cytosolic DNA, whereas mutations that affect the dNTPase activity of SAMHD1 do not, confirming thereby that defective fork processing and aberrant induction of the IFN pathway are directly linked. This mechanism is distinct from the induction of cGAS–STING by micronuclei recently reported in cells deficient for RNase H2, the enzyme encoded by the *RNASEH2A*, *RNASEH2B* and *RNASEH2C* genes, which are also frequently mutated in AGS^{38,39}. Because replication stress contributes to cancer development¹⁸, our findings also shed new light on the mechanism by which SAMHD1 mutations might promote tumorigenesis.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0050-1>.

Received: 30 May 2017; Accepted: 7 March 2018;

Published online: 18 April 2018

- Goldstone, D. C. et al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379–382 (2011).
- Laguette, N. et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654–657 (2011).
- Hrecka, K. et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658–661 (2011).
- White, T. E. et al. The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation. *Cell Host Microbe* **13**, 441–451 (2013).
- Cribier, A., Descours, B., Valadão, A. L., Laguette, N. & Benkirane, M. Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. *Cell Reports* **3**, 1036–1043 (2013).
- Crow, Y. J. & Manel, N. Aicardi–Goutières syndrome and the type I interferonopathies. *Nat. Rev. Immunol.* **15**, 429–440 (2015).
- Yang, Y.-G., Lindahl, T. & Barnes, D. E. Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. *Cell* **131**, 873–886 (2007).
- Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP–AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
- Maelfait, J., Bridgeman, A., Benlahrech, A., Cursi, C. & Rehwinkel, J. Restriction by SAMHD1 limits cGAS/STING-dependent innate and adaptive immune responses to HIV-1. *Cell Reports* **16**, 1492–1501 (2016).
- Zhao, K. et al. Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi–Goutières syndrome-related SAMHD1. *Cell Reports* **4**, 1108–1115 (2013).
- Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R. & Chédin, F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi–Goutières syndrome. *eLife* **4**, e08007 (2015).
- Seamon, K. J., Sun, Z., Shlyakhtenko, L. S., Lyubchenko, Y. L. & Stivers, J. T. SAMHD1 is a single-stranded nucleic acid binding protein with no active site-associated nuclease activity. *Nucleic Acids Res.* **43**, 6486–6499 (2015).
- Beloglazova, N. et al. Nuclease activity of the human SAMHD1 protein implicated in the Aicardi–Goutières syndrome and HIV-1 restriction. *J. Biol. Chem.* **288**, 8101–8110 (2013).
- Ryoo, J. et al. The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat. Med.* **20**, 936–941 (2014).

15. Clifford, R. et al. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* **123**, 1021–1031 (2014).
16. Rentoft, M. et al. Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proc. Natl Acad. Sci. USA* **113**, 4723–4728 (2016).
17. Franzolin, E. et al. The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc. Natl Acad. Sci. USA* **110**, 14272–14277 (2013).
18. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).
19. Pasero, P. & Vindigni, A. Nucleases acting at stalled forks: how to reboot the replication program with a few shortcuts. *Annu. Rev. Genet.* **51**, 477–499 (2017).
20. Hashimoto, Y., Ray Chaudhuri, A., Lopes, M. & Costanzo, V. Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat. Struct. Mol. Biol.* **17**, 1305–1311 (2010).
21. Schlacher, K. et al. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529–542 (2011).
22. Ray Chaudhuri, A. et al. Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature* **535**, 382–387 (2016).
23. Langereis, M. A., Rabouw, H. H., Holwerda, M., Visser, L. J. & van Kuppeveld, F. J. M. Knockout of cGAS and STING rescues virus infection of plasmid DNA-transfected cells. *J. Virol.* **89**, 11169–11173 (2015).
24. Holm, C. K. et al. Influenza A virus targets a cGAS-independent STING pathway that controls enveloped RNA viruses. *Nat. Commun.* **7**, 10680 (2016).
25. Sirbu, B. M. et al. Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes Dev.* **25**, 1320–1327 (2011).
26. Zellweger, R. et al. Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J. Cell Biol.* **208**, 563–579 (2015).
27. Bétous, R. et al. SMARCA1 catalyzes fork regression and Holliday junction migration to maintain genome stability during DNA replication. *Genes Dev.* **26**, 151–162 (2012).
28. Kolinjivadi, A. M. et al. Smarcal1-mediated fork reversal triggers mre11-dependent degradation of nascent DNA in the absence of Brca2 and stable Rad51 nucleofilaments. *Mol. Cell* **67**, 867–881.e7 (2017).
29. Cruz-García, A., López-Saavedra, A. & Huertas, P. BRCA1 accelerates CtIP-mediated DNA-end resection. *Cell Reports* **9**, 451–459 (2014).
30. Gunn, A., Bennardo, N., Cheng, A. & Stark, J. M. Correct end use during end joining of multiple chromosomal double strand breaks is influenced by repair protein RAD50, DNA-dependent protein kinase DNA-PKcs, and transcription context. *J. Biol. Chem.* **286**, 42470–42482 (2011).
31. Daddacha, W. et al. SAMHD1 promotes DNA end resection to facilitate dna repair by homologous recombination. *Cell Reports* **20**, 1921–1935 (2017).
32. Lee, J. & Dunphy, W. G. The Mre11-Rad50-Nbs1 (MRN) complex has a specific role in the activation of Chk1 in response to stalled replication forks. *Mol. Biol. Cell* **24**, 1343–1353 (2013).
33. Jazayeri, A., Balestrini, A., Garner, E., Haber, J. E. & Costanzo, V. Mre11-Rad50-Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J.* **27**, 1953–1962 (2008).
34. Härtlova, A. et al. DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti-microbial innate immunity. *Immunity* **42**, 332–343 (2015).
35. Shen, Y. J. et al. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Reports* **11**, 460–473 (2015).
36. Ho, S. S. W. et al. The DNA structure-specific endonuclease MUS81 mediates DNA sensor STING-dependent host rejection of prostate cancer cells. *Immunity* **44**, 1177–1189 (2016).
37. Wolf, C. et al. RPA and Rad51 constitute a cell intrinsic mechanism to protect the cytosol from self DNA. *Nat. Commun.* **7**, 11752 (2016).
38. Mackenzie, K. J. et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).
39. Bartsch, K. et al. Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. *Hum. Mol. Genet.* **26**, 3960–3972 (2017).

Acknowledgements We thank N. Laguet, G. Uzé, F. van Kuppeveld and S. Paludan for reagents and cell lines. We also thank J. Cau and the MRI Imaging facility for help with image analysis and the DNA combing facility of Montpellier for silanized coverslips. We thank A. Constantinou, D. Maiorano, B. Pardo, M. Moriel-Carretero and H. Tourrière for discussions and C. Featherstone for professional editing. This work was supported by grants from the Agence Nationale pour la Recherche (ANR) and Institut National du Cancer (INCa) to P.P. and B.L., the Ligue Contre le Cancer (équipe labellisée), Cancerpole Grand Sud-ouest and SIRIC Montpellier Cancer to P.P. and the MSDAvenir fund to P.P. and M.B. Work in W.N.'s laboratory is supported by funding from the Medical Research Council and the Institute of Cancer Research. Work in A.C.'s laboratory is supported by the Swedish Cancer Society and the Swedish Research Council. Work in L.K.'s laboratory is supported by the Czech Science Foundation grants (GACR 17-17720S and 13-26629S) and project no. LQ1605 from the National Program of Sustainability II (MEYS CR). M.J.S. was supported by fellowships from CNRS, Region LR and Fondation ARC. F.C. is supported by a fellowship from the French MRES and from Fondation Recherche Médicale. This work was also funded by the Associazione Italiana per Ricerca sul Cancro (AIRC), the European Research Council (ERC) Consolidator Grant 614541, an Armenise-Harvard Foundation career development award and an AICR Worldwide Cancer Research award (13-0026) to V.C.

Reviewer information Nature thanks L. Zou and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.-L.L. and P.P. conceived and planned the study. Y.-L.L., F.C., M.-J.S., A.-L.S. and A.P. designed and performed experiments on replication stress. S.S. and A.Ch. analysed dNTP pools. C.M. and A.Cr. constructed SAMHD1 mutants. Y.-L.L., A.B. and C.M. performed qRT-qPCR experiments. J.N. and W.N. performed iPOND experiments. E.D. and B.L. performed SSA assay. A.S. helped with immunofluorescence experiments. V.C. designed and H.T. performed experiments with *Xenopus* Samhd1. K.Z. and L.K. designed, performed and analysed nuclease assays. Y.-L.L., M.B. and P.P. wrote the manuscript and all authors reviewed it.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0050-1>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0050-1>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to Y.-L.L. or P.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and cell culture. HEK293T cells (ATCC CRL-3216) and SAMHD1-overexpressing HeLa cells¹⁵ were cultured at 37 °C under 5% CO₂ in DMEM supplemented with 10% fetal calf serum (FCS). Epstein–Barr virus-immortalized B cells from a *SAMHD1*^{-/-} patient were cultured in RPMI-1640 supplemented with 10% FCS. Reporter HL116 cells were cultured in DMEM supplemented with 10% FCS and ultra-glutamine under HAT selection⁴⁰.

Construction of SAMHD1 mutants. A SAMHD1 wild-type clone (Invitrogen; clone ID: IOH55544, sequence NM 015474.3) with a C-terminal Flag and a haemagglutinin (HA) tag was modified to produce a shRNA-resistant coding sequence. The DNA was amplified by PCR with the appropriate primers to create a SAMHD1 protein-coding sequence with an HA tag fused to the second amino acid residue from the N terminus and to restore the normal C terminus. The PCR reaction also created a MluI site after the sequence encoding the protein C terminus and an AgeI site followed by a Kozac sequence before the sequence encoding the protein N terminus. The amplified DNA was cloned in p-TripZ (Thermo Scientific Open Biosystems) between AgeI and MluI. All mutations were created by using the QuickChange IIXL kit from Agilent. The complete SAMHD1-coding sequence was checked by sequencing for each mutant. K312A: AAA to GCA; Y315A: TAT to GCT; T592A: ACA to GCA; T592E: ACA to GAA. HIV-1-derived lentiviral vectors were produced in HEK293T cells as previously described⁴¹.

Detection of cytosolic ssDNA. Cells were treated with 4 mM HU for 2 h, or untreated, as appropriate, and then incubated in fresh medium for 2 h and fixed for 20 min with 80% methanol at -20 °C. Coverslips were incubated with a mouse anti-ssDNA antibody (see Supplementary Information) overnight at 4 °C and with a secondary antibody conjugated to an Alexa Fluor dye for 1 h at room temperature, followed by DAPI staining. Images were acquired by using a Zeiss ApoTome microscope and a LSM780 confocal microscope. For the detection of BrdU-labelled cytosolic DNA, cells were labelled for 2 h with 10 µM BrdU before HU exposure and BrdU was detected with a rat anti-BrdU antibody without denaturation. The mean fluorescence intensity of cytosolic ssDNA and BrdU was quantified with the CellProfiler software. Nuclei were segmented using DAPI signal and masked to quantify ssDNA and BrdU signals. Statistical analysis was performed with GraphPad Prism (GraphPad Software) using two-sided Mann–Whitney rank sum test.

Quantification of cytosolic mitochondrial DNA. Cells growing on 100-mm coverslips were collected into 200 µl fractionation buffer (20 mM HEPES, pH 7.4, 10 mM KCl, 2 mM MgCl₂, 1 mM EDTA, 1 mM EGTA, 1 mM DTT) and left on ice for 15 min. They were then lysed by passing cell suspensions 10 times through a 27-gauge needle attached to a 1-ml syringe and left on ice for a further 20 min. Cell lysates were centrifuged at 720g for 5 min at 4 °C. The supernatant was further centrifuged at 10,000g for 5 min at 4 °C to remove mitochondria. Cytosolic DNA was purified from the supernatant by using DNeasy Blood & Tissue kit (Qiagen) according to the manufacturer's instructions. The presence in the cytosolic fraction of the mitochondrial gene *MT-CO1* encoding the cytochrome c oxidase 1 (COX1) was quantified by qPCR using specific primers.

Analysis of IFN production. Cell culture media were collected 48 h after transfection with various siRNAs. Reporter HL116 cells harbouring a type-I IFN-inducible 6-16 promoter fused to the firefly luciferase gene⁴⁰ were cultured for 48 h in the presence of the collected culture media. Luciferase activity was quantified according to the manufacturer's instructions (Luciferase assay system, Promega). The expression of various IFNs and IFN-stimulated genes (ISGs) was quantified by qRT-PCR (see Supplementary Information) and normalized to *GAPDH*. Reverse transcription was performed by using the Superscript III First Synthesis System for RT-PCR (ref. 18080-051, Invitrogen). A LightCycler 480 SYBR Green I Master Mix (ref. 04887352001, Invitrogen) was used to perform quantitative PCR.

Cell sorting. HEK293T cells expressing the Scramble-control shRNA (shScr) were labelled with propidium iodide. Populations of G₁, S and G₂/M phase cells were sorted in a FACSAria sorter (Becton Dickinson, MRI facility).

Flow cytometry analysis of S phase progression. HEK293T cells transduced with lentiviral vectors expressing shScr or shSAMHD1 (shSAM) RNAs were pulse labelled with 10 µM EdU for 15 min and then chased with 100 µM thymidine for the indicated periods of time. After fixation with 1% formaldehyde for 30 min at room temperature, EdU incorporation was detected by using Click chemistry according to the manufacturer's instructions (Click-iT EdU Flow Cytometry Cell Proliferation Assay, Invitrogen). The cells were resuspended in PBS containing 1% (w/v) BSA, 2 µg ml⁻¹ DAPI and 0.5 mg ml⁻¹ RNase A for 30 min at room temperature and were analysed in a MACSQuant flow cytometer (Miltenyi Biotec). The percentages of cells in G₁, S and G₂/M phases were quantified by using FlowJo single-cell analysis software (FlowJo, LLC).

Confocal microscopy of DNA replication foci. HeLa cells overexpressing HA-tagged SAMHD1 and GFP proteins were grown on coverslips. They were labelled for 1 h with 10 µM EdU then fixed with 1% paraformaldehyde (PFA) and permeabilized with 0.1% saponin. Coverslips were then incubated overnight at 4 °C with an anti-HA antibody. Replication foci were detected by using Click chemistry, as for flow cytometry analysis. The cell nuclei were stained with DAPI and the coverslips were then mounted on glass slides and visualized by using a Zeiss LSM780 confocal microscope. The percentage of HA–EdU foci colocalization was quantified by ImageJ-JACoP (imagej.nih.gov/ij/plugins/track/jacop.html). Statistical analysis was performed with GraphPad Prism (GraphPad Software) using two-sided Mann–Whitney rank sum test.

iPOND. HeLa S3 cells (1 × 10⁶ per ml) or HEK293FT cells growing logarithmically were incubated with 10 µM EdU for 10 min. After EdU labelling, cells were fixed in 1% formaldehyde, quenched by adding glycine to a final concentration of 0.125 M and washed three times in PBS. Collected cell pellets were frozen at -80 °C and cells were permeabilized by resuspending 1.0–1.5 × 10⁷ cells per ml in ice-cold 0.25% Triton X-100 in PBS and incubating for 30 min. Before the Click reaction, samples were washed once in PBS containing 0.5% BSA and once in PBS. Cells were incubated for 1 h at room temperature in Click reaction buffer containing 10 µM azide-PEG(3+3)-S-S-biotin conjugate (Click ChemistryTools, cat. no AZ112-25), 10 mM sodium ascorbate, and 1.5 mM CuSO₄ in PBS. The 'no Click' reaction contained DMSO instead of biotin-azide. Following the Click reaction, cells were washed once in PBS containing 0.5% BSA and once in PBS. Cells were resuspended in lysis buffer (50 mM Tris–HCl pH 8.0, 1% SDS) containing protease inhibitor cocktail (Sigma) and sonicated with a Diagenode Bioruptor Plus for 40 cycles (30 s on/30 s off). Samples were centrifuged at 14,500g at 4 °C for 30 min and the supernatant was diluted 1:3 with TNT buffer (50 mM Tris pH 7.5, 200 mM NaCl, 0.3% Triton X-100) containing protease inhibitors. An aliquot was taken as an input sample. Streptavidin–agarose beads (Novagen) were washed three times in TNT buffer containing protease inhibitor cocktail. Two hundred microlitres of bead slurry were used per 1 × 10⁸ cells. The streptavidin–agarose beads were resuspended 1:1 in TNT buffer containing protease inhibitors and added to the samples, which were then incubated at 4 °C for 16 h in the dark. Following binding, the beads were washed twice with 1 ml TNT buffer, twice with TNT buffer containing 1 M NaCl and twice with TNT buffer. Protein–DNA complexes were eluted by incubating with 5 mM DTT in TNT buffer. Cross-links were reversed by incubating samples in SDS sample buffer at 95 °C for 20 min. Proteins were resolved on SDS–PAGE and detected by immunoblotting using specific antibodies.

DNA fibre spreading. DNA fibre spreading was performed as described previously⁴². In brief, subconfluent cells were sequentially labelled first with 10 µM 5-iodo-2'-deoxyuridine (IdU) and then with 100 µM 5-chloro-2'-deoxyuridine (CldU) for the indicated times. One thousand cells were loaded onto a glass slide (StarFrost) and lysed with spreading buffer (200 mM Tris–HCl pH 7.5, 50 mM EDTA, 0.5% SDS) by gently stirring with a pipette tip. The slides were tilted slightly and the surface tension of the drops was disrupted with a pipette tip. The drops were allowed to run down the slides slowly, then air dried, fixed in 3:1 methanol:acetic acid for 10 min, and allowed to dry. Glass slides were processed for immunostaining with mouse anti-BrdU to detect IdU, rat anti-BrdU to detect CldU, mouse anti-ssDNA antibodies (see Supplementary Information for details) and corresponding secondary antibodies conjugated to various Alexa Fluor dyes. Nascent DNA fibres were visualized by using immunofluorescence microscopy (Leica DM6000 or Zeiss ApoTome). The acquired DNA fibre images were analysed by using MetaMorph Microscopy Automation and Image Analysis Software (Molecular Devices). Statistical analysis was performed with GraphPad Prism (GraphPad Software) using two-sided Mann–Whitney rank sum test.

SMART. Control (shScr) and SAMHD1-depleted (shSAM) HEK293T cells were labelled with 10 µM BrdU for 24 h. They were then treated with 5 µM bleocin (Calbiochem) for 1 h and collected at the indicated time points. They were processed for DNA fibre spreading as described. BrdU tracks were stained with anti-BrdU antibody without DNA denaturation and visualized by fluorescence microscopy (Zeiss ApoTome). The acquired DNA fibre images were analysed by using MetaMorph Microscopy Automation and Image Analysis Software (Molecular Devices) and statistical analysis was performed with GraphPad Prism (GraphPad Software) using two-sided Mann–Whitney rank sum test.

Single-strand annealing assay. Seven hours after plating in six-well plates (1 × 10⁵ cells per well), U2OS single-strand annealing cells were transfected with siRNAs against *SAMHD1* or *ChIP*, or with control siRNAs (see Supplementary Information for sequences) by using interferin reagent (Polyplus, Ozyme). Forty-eight hours after siRNA transfection with siRNAs, HA-tagged I-SceI was expressed by transient transfection (JetPEI, Polyplus, Ozyme) with 1.5 µg of the expression plasmid pCMV-I-SceI⁴³ for DSB induction. Expression was verified by immunoblotting with an anti-HA antibody. Forty-eight hours after transfection with the construct expressing I-SceI, cells were fixed in PBS containing 2% formaldehyde for 15 min

at room temperature. Cells expressing GFP were monitored by flow cytometry in a BD Accuri C6 flow cytometer.

Collection and processing of HEK293T cells for nucleotide pool measurements. Three 150 × 20 mm cell culture dishes were seeded with (8×10^6) HEK293T cells and grown to 80% confluency. After removal of the medium, the plates were washed twice with cold NaCl (9 g l^{-1}). The cells were scraped with a cell scraper after addition of 250 μl of cold 15% TCA containing 30 mM MgCl_2 to each of the three plates and they were pooled together. The pooled cells were collected in an Eppendorf tube and snap-frozen in liquid nitrogen. They were stored at -80°C until processed. For extraction of dNTPs and NTPs, cells were thawed on ice and vortex-mixed for 10 min in a cold room. The supernatant was then collected by centrifugation at 14,000g for 5 min at 4°C and processed as described⁴⁴.

Detection of ssDNA, RPA and MRE11 foci. Cells growing on coverslips were labelled with 10 μM BrdU for 24 h. They were treated with 1 mM HU for 2 h after the removal of BrdU and fixed with ice-cold methanol for 1 h at -20°C . The coverslips were incubated with an anti-BrdU monoclonal antibody (without DNA denaturation) overnight at 4°C and then with a secondary antibody conjugated to an Alexa Fluor dye for 1 h at room temperature, followed by DAPI staining. Images were acquired with a wide-field Leica DM6000 microscope and a Zeiss LSM780 confocal microscope. The number of subnuclear ssDNA foci was quantified by using CellProfiler image analysis software. For detection of chromatin-bound RPA foci during S phase, cells growing on coverslips were pulse labelled with 10 μM EdU for 10 min and treated with 4 mM HU for 2 h. They were fixed with 4% PFA in PBS for 15 min and then incubated for 3 min at 4°C with CSK buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1 mM MgCl_2 , 1 mM EGTA, 300 mM sucrose, 0.5 mM DTT) containing 0.25% Triton X-100 and phosphatase inhibitor cocktail (Sigma-Aldrich, P0044). EdU incorporation was detected by using Click chemistry and then blocked with 3% BSA in PBS for 1 h at room temperature. The coverslips were incubated with an anti-RPA antibody (overnight at 4°C) and then with a secondary antibody conjugated to an Alexa Fluor dye for 1 h at 37°C , followed by DAPI staining. Images were acquired by using a Zeiss ApoTome microscope. The mean fluorescence intensity in EdU-positive cells was quantified by using CellProfiler (<http://www.cellprofiler.org>). For the detection of MRE11 foci, cells seeded on coverslips were labelled with EdU and treated with or without HU as described above. They were incubated in cold extraction buffer (20 mM HEPES pH 7.5, 50 mM NaCl, 300 mM sucrose, 3 mM MgCl_2 , 0.5% Triton X-100) for 5 min at 4°C and then fixed in fixation buffer (3.7% PFA, 2% sucrose, 0.5% Triton X-100) for 20 min at room temperature. The coverslips were incubated with an anti-MRE11 antibody overnight at 4°C after blocking in PBS containing 1% BSA in PBS for 1 h at room temperature. After incubation with a secondary antibody conjugated to Alexa Fluor dye, EdU incorporation was detected by using Click chemistry, followed by DAPI staining. Images were acquired by using a Zeiss LSM780 confocal microscope. The percentage of colocalized MRE11 and EdU foci was quantified by ImageJ-JACOPI. **Cloning and purification of recombinant *Xenopus* Samhd1.** The reference sequence of *Xenopus* Samhd1 (xSamhd1), available from Uniprot, Q6INN8 (<http://www.uniprot.org>), contains 632 amino acids. In brief, the cDNA encoding *Xenopus laevis* samhd1 (XGC Samhd1, clone ID 3402629, from Thermo Fisher Scientific) was cloned into the pET43 plasmid following standard methods. Clones were validated by PCR and DNA sequencing.

Purification of recombinant xSamhd1 (His- or His-Flag-tagged) was performed in two steps after IPTG (0.5 mM) induction of the bacteria overnight at 23°C . The bacterial pellet was lysed in a buffer containing 500 mM NaCl, 50 mM HEPES pH 7.5, 2 mM β -mercaptoethanol, 10% glycerol and protease inhibitors cocktail, then sonicated and clarified by centrifugation. The first purification step was performed by binding the recombinant xSamhd1 to Ni-NTA resin (HisPur, Thermo Fisher Scientific) for 2 h at 4°C and eluting with 100 mM and 250 mM imidazole. After dialysis, the recombinant protein was further purified by size exclusion chromatography (Superdex 200 10/300 GL, GE healthcare). For use in *Xenopus* egg extracts, both His- and His-Flag- xSamhd1 were purified in buffer containing 150 mM NaCl, 25 mM Tris-HCl pH 7.6, 1 mM DTT, 10% glycerol.

Production of anti-xSamhd1 antibody. For antibody production, the recombinant His-xSamhd1 was purified, as described above, but size exclusion chromatography was performed in phosphate-buffered saline. This preparation was supplied to BioGenes GmbH, who immunised rabbits and prepared the antiserum. The antibody was affinity purified from the antiserum by binding to recombinant His-xSamhd1 by the Cogentech service of the IFOM protein facility.

Preparation of interphase *Xenopus* egg extracts and treatments. *Xenopus* egg extracts were prepared as described⁴⁵. The DNA damage response and DNA replication were analysed as previously described^{45,46}. To induce DNA double-strand breaks, we added $0.05 \text{ U } \mu\text{l}^{-1}$ EcoRI (New England Biolabs) to the extracts. To slow or stall replication forks, we added aphidicolin (Sigma) to the extracts at the concentrations indicated in the main text.

Immunodepletion of xSamhd1 in *Xenopus* egg extracts. To immunodeplete xSamhd1 from *Xenopus* egg extracts, 0.3–0.5 ml of egg extract were incubated with

affinity-purified IgGs (30 mg) and with 120 μl of Dynabeads-ProteinA (Thermo Fisher, 10002D). The immunodepletion was performed in three rounds of 30–45 min at 4°C .

Immunoprecipitation of xSamhd1 in *Xenopus* egg extract. For coimmunoprecipitation of xSamhd1 with RPA70, 200 μl of Dynabeads-Protein A slurry were coupled with 60 μg of anti-xSamhd1 purified serum or 50 μg of rabbit pre-immune serum (IgG control). The beads were then incubated with 500 μl of clarified *Xenopus* egg extract brought to a final volume of 1.5 ml with NIB buffer (50 mM KCl, 50 mM HEPES pH 7.0, 5 mM MgCl_2 , 2 mM DTT, 0.5 mM spermidine, 0.15 mM spermine, 20% sucrose, 0.05% Triton X-100). After 1 h at 4°C , a 'flow-through' fraction of the unbound material was recovered and boiled in Laemmli sample buffer (Biorad) for further analysis. Beads were washed five times in NIB buffer and boiled in Laemmli sample buffer before analysis by western blotting. For coimmunoprecipitation of xSamhd1 with xCtIP, 100 μl Dynabeads-Protein A slurry was coupled to 30 μg anti-xSamhd1 (affinity purified serum) or 50 μg of pre-immune serum (IgG control). The beads were then incubated with 300 μl of clarified *Xenopus* egg extract brought to a final volume of 1 ml with a buffer containing 100 mM KCl, 50 mM HEPES pH 7.0, 5 mM MgCl_2 , 2 mM DTT, 20% sucrose, 0.05% Triton X-100. After overnight incubation at 4°C , a fraction of the flow-through was recovered and boiled in Laemmli sample buffer for further analysis. Beads were washed five times in the immunoprecipitation buffer and boiled in Laemmli sample buffer before analysis by western blotting.

Biotin-dUTP pulldown of nascent chromatin. Biotin-dUTP pulldown was performed as previously described²⁸. In brief, *Xenopus* sperm nuclei were added to 100 μl *Xenopus* egg extracts at a final concentration of 4,000 nuclei per μl . Forty-five minutes after sperm nuclei addition, newly synthesized DNA in the extracts was labelled with 40 μM biotin-16-dUTP (Roche) in the presence of either 5 μM aphidicolin or DMSO as control for 10 min. DNA replication was stopped by the addition of 200 μl cold EB-EDTA buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 2.5 mM MgCl_2 , 1 mM EDTA). Samples were homogenized by using a cut Eppendorf p1000 pipette tip and overlaid on 600 μl EB-EDTA buffer containing 30% (w/v) sucrose. Nuclei were collected by centrifugation at 8,300g at 4°C , the nuclear pellets were resuspended in 400 μl EB-NP40 buffer (50 mM HEPES-KOH pH 7.5, 100 mM KCl, 2.5 mM MgCl_2 , 0.25% NP40) and then subjected twice to 10 min sonication with a Bioruptor device set to maximum power. After the sonication step, 20 μl from each sample were set aside (5% input). Biotinylated DNA fragments were then pulled-down by incubation with 40 μl Dynabeads M-280 Streptavidin for 1 h at 4°C . The beads were then washed three times in EB-EDTA buffer and boiled with 30 μl of $1\times$ denaturing loading buffer, and the entire volume was separated by SDS-PAGE and analysed by western blotting.

Protein purification. 10 \times His-SAMHD1 was overexpressed from pET19b as previously described⁴⁷. All purification steps were carried out at 4°C . The cell pellet was resuspended and lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM β -mercaptoethanol, 1 mM EDTA, protease inhibitors) and centrifuged for 1 h at 35,000g. Soluble extract was incubated with pre-equilibrated Ni-NTA agarose resin (Qiagen) for 2 h with continuous mixing. The resin was washed with buffer T (20 mM Tris-HCl pH 7.5, 10% glycerol, 1 mM EDTA) and bound protein was eluted with a gradient of 100–500 mM imidazole in buffer T. The eluted fractions were diluted in buffer T, loaded onto a 1 ml SourceS column and the protein was eluted with a 10-ml gradient of 100–700 mM KCl in buffer T. The peak fractions were stored in small aliquots at -80°C .

MRE11-6 \times His and yMRE11 were purified from Sf9 insect cells or yeast cells, respectively, as described previously⁴⁸. DNA2 was purified from Sf9 cells as described previously⁴⁹ and FEN1 was expressed and purified by the same protocol as described for yFEN1⁵⁰. Bacterial ExoIII was purchased from Thermo Fisher Scientific (EN0191).

Nuclease assays. Fluorescently labelled DNA substrates used for nuclease assays were prepared by annealing as described elsewhere⁵¹. The concentrations of SAMHD1 indicated in the main text were pre-incubated for 5 min at 37°C with MRE11 (4 nM) or yMRE11 (4 nM) followed by addition of a DNA substrate and incubation in buffer P (40 mM KCl, 10 mM Tris-HCl (pH 7.5), 1 mM DTT, 10 $\mu\text{g ml}^{-1}$ BSA) in the presence of 2 mM dGTP, 2 mM MnCl_2 and 2 mM MgCl_2 for 50 min at 37°C . The reaction mixtures were then incubated with 0.1% SDS and 500 $\mu\text{g ml}^{-1}$ proteinase K at 37°C for 20 min, heat-denatured, the DNA was resolved on 20% denaturing PAGE gels (acrylamide:bisacrylamide, 19:1) and scanned with a Fuji FLA 9000 imager. Where indicated, gels were quantified using Multi Gauge v.3.2 software (Fuji). For nuclease assays with DNA2, reactions were carried out as for MRE11 and yMRE11 but using 0.4 nM DNA2 in buffer P containing 2 mM ATP, 2 mM MgCl_2 and 2 mM dGTP. Assays of FEN1 (1 nM) and ExoIII (20 $\mu\text{U } \mu\text{l}^{-1}$) were carried out in buffer P containing 2 mM MgCl_2 and 2 mM dGTP.

Electrophoretic mobility shift assays. Fluorescently labelled DNA substrates were prepared as for the nuclease assays⁵¹. SAMHD1 was incubated with fluorescently

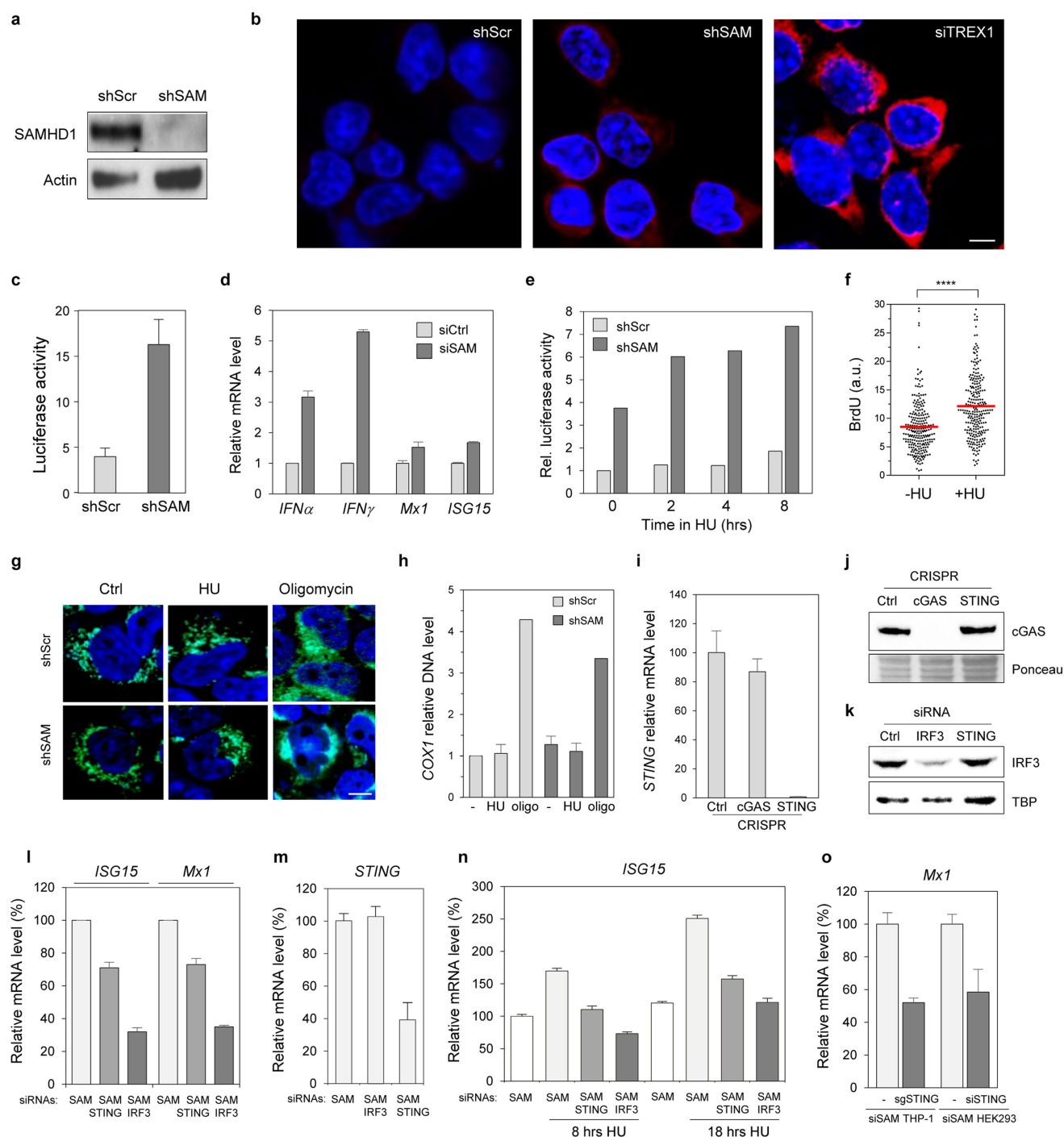
labelled DNA substrate for 10 min at 37 °C in the presence of 30 mM KCl, 10 mM Tris-HCl pH 7.5, 1 mM DTT, 2 mM MgCl₂. Protein–DNA complexes were cross-linked by incubation with 0.01% glutaraldehyde for 5 min at 37 °C and resolved on 0.8% agarose gels. The gels were scanned with a Fuji FLA 9000 imager and quantified using Multi Gauge v.3.2 software (Fuji).

Microscale thermophoresis. For MRE11–SAMHD1 interaction measurements, MRE11 was fluorescently labelled by using Monolith Protein Labelling Kit RED-MALEIMIDE (Cysteine Reactive) according to the manufacturer's protocol. Measurements were performed with 8 nM labelled MRE11 and the indicated concentration of SAMHD1 in PBS containing 0.05% Tween. Thermophoresis was performed on a Monolith NT.115 (NanoTemper Technologies GmbH) set at 50% LED and 80% MST power at 25 °C and with 5 s and 30 s laser off and on times, respectively. For RPA–SAMHD1 interaction measurements, 10×His–SAMHD1 was labelled by using the Monolith His-Tag Labelling Kit RED-tris-NTA according to the manufacturer's protocol. The thermophoresis measurements were performed with 50 nM SAMHD1 and increasing concentration of RPA in PBS containing 0.05% Tween and 2 mM MgCl₂ a Monolith NT.115 set at 40% LED and 80% MST power at 25 °C and with 5 s and 30 s laser off and on times, respectively. Data were analysed by using MO.Affinity Analysis Software v.2.2.4 (NanoTemper Technologies GmbH).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The authors declare that the data supporting the findings of this study are available within the Article and its Supplementary Information files or are available from corresponding authors.

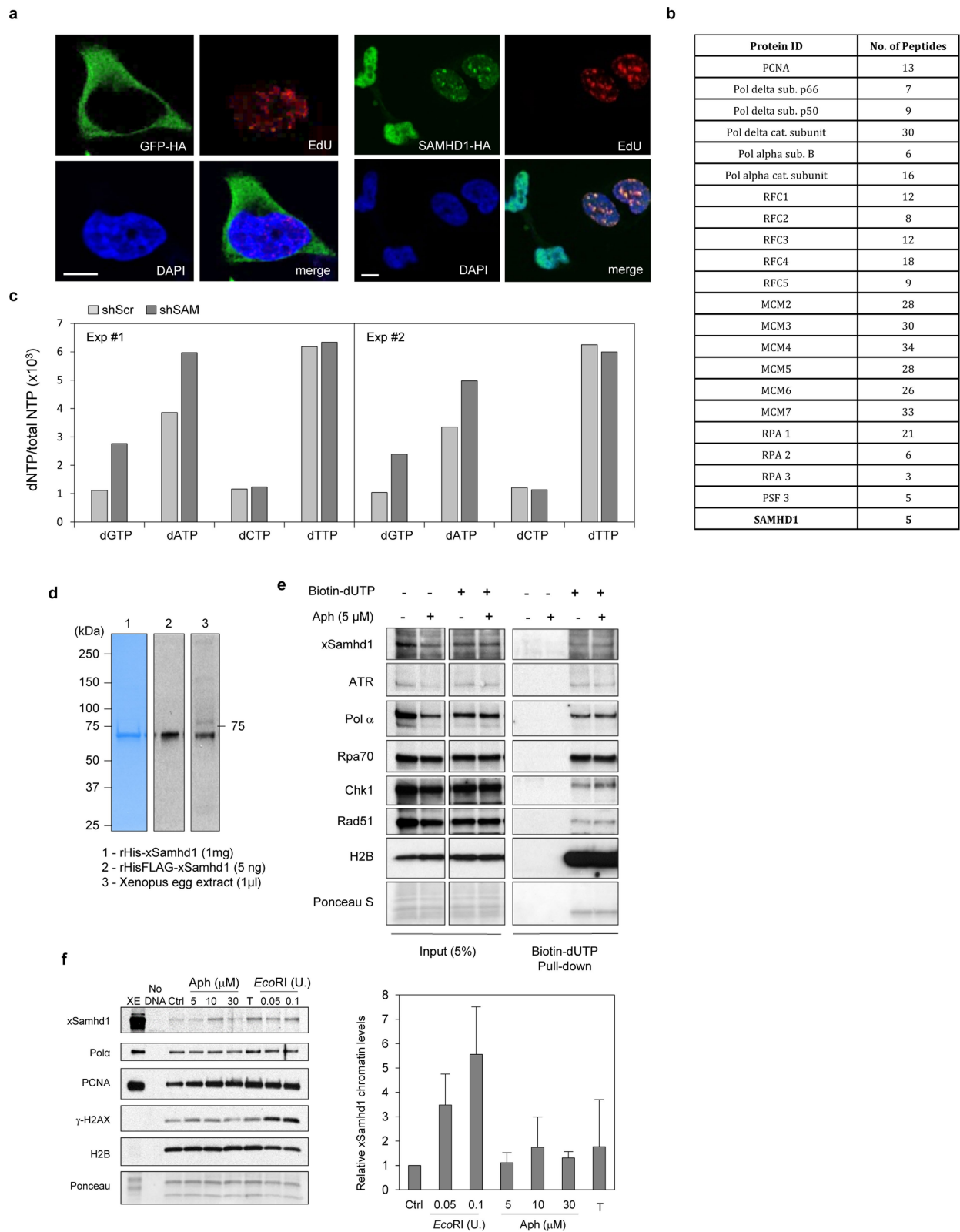
40. Garcin, G. et al. High efficiency cell-specific targeting of cytokine activity. *Nat. Commun.* **5**, 3016 (2014).
41. Lin, Y. L. et al. Feline immunodeficiency virus vectors for efficient transduction of primary human synoviocytes: application to an original model of rheumatoid arthritis. *Hum. Gene Ther.* **15**, 588–596 (2004).
42. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295 (1998).
43. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl Acad. Sci. USA* **95**, 5172–5177 (1998).
44. Jia, S., Marjavaara, L., Buckland, R., Sharma, S. & Chabes, A. Determination of deoxyribonucleoside triphosphate concentrations in yeast cells by strong anion-exchange high-performance liquid chromatography coupled with ultraviolet detection. *Methods Mol. Biol.* **1300**, 113–121 (2015).
45. Sannino, V., Pezzimenti, F., Bertora, S. & Costanzo, V. *Xenopus laevis* as model system to study DNA damage response and replication fork stability. *Methods Enzymol.* **591**, 211–232 (2017).
46. Hashimoto, Y. & Costanzo, V. Studying DNA replication fork stability in *Xenopus* egg extract. *Methods Mol. Biol.* **745**, 437–445 (2011).
47. Hansen, E. C., Seamon, K. J., Cravens, S. L. & Stivers, J. T. GTP activator and dNTP substrates of HIV-1 restriction factor SAMHD1 generate a long-lived activated state. *Proc. Natl Acad. Sci. USA* **111**, E1843–E1851 (2014).
48. Zadorozhny, K. et al. Fanconi-anemia-associated mutations destabilize Rad51 filaments and impair replication fork protection. *Cell Reports* **21**, 333–340 (2017).
49. Pinto, C., Kasaciunaite, K., Seidel, R. & Cejka, P. Human DNA2 possesses a cryptic DNA unwinding activity that functionally integrates with BLM or WRN helicases. *eLife* **5**, e18574 (2016).
50. Matulova, P. et al. Cooperativity of Mus81.Mms4 with Rad54 in the resolution of recombination and replication intermediates. *J. Biol. Chem.* **284**, 7733–7745 (2009).
51. Marini, V. & Krejci, L. Unwinding of synthetic replication and recombination substrates by Srs2. *DNA Repair (Amst.)* **11**, 789–798 (2012).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | SAMHD1-depleted cells secrete IFNs. **a**, Western blot analysis of SAMHD1 in HEK293T cells expressing *SAMHD1* shRNA (shSAM) or a scrambled control (shScr). **b**, Cytosolic ssDNA (red) in shScr and shSAM HEK293T cells and in HEK293T cells transfected with *TREX1* siRNA (siTRESX1). Scale bar, 5 μ m. $n = 5$. **c**, HL116 cells containing an IFN-stimulated response element–luciferase reporter gene were incubated with culture medium from shScr or shSAM HEK293T cells for 48 h. Mean luciferase activity and s.d. from four independent experiments are shown. **d**, HeLa cells were transfected for 48 h with control siRNA (siCtrl) or *SAMHD1* siRNA (siSAM). Expression of *IFNA*, *IFNG* and the IFN-stimulated genes *MX1* and *ISG15* was quantified by qRT–PCR. Data are representative of three independent experiments. Error bars denote s.d. for a representative triplicate experiment. **e**, shScr or shSAM HEK293T cells were treated with 4 mM HU for the indicated times and then transferred to fresh medium for a total of 20 h. Culture medium was collected and incubated with HL116 cells for 48 h before luciferase assay. Data shown are representative of three independent experiments. **f**, Quantification of the mean fluorescence intensity of cytosolic BrdU in the experiment shown in Fig. 1e. Quantification was performed on 250 cells by using CellProfiler. Median BrdU intensity is indicated in red. **** $P < 0.0001$, Mann–Whitney rank sum test. **g**, shScr and shSAM HEK293T cells were treated for 2 h with 4 mM HU or for 6 h with 40 μ M oligomycin, used here to damage mitochondria. Cells were labelled with the mitochondria-selective dye MitoTracker (Invitrogen). The integrity of mitochondria was assessed by confocal microscopy. Representative images are shown. Scale bar, 5 μ m. **h**, The abundance of mitochondrial *COX1*

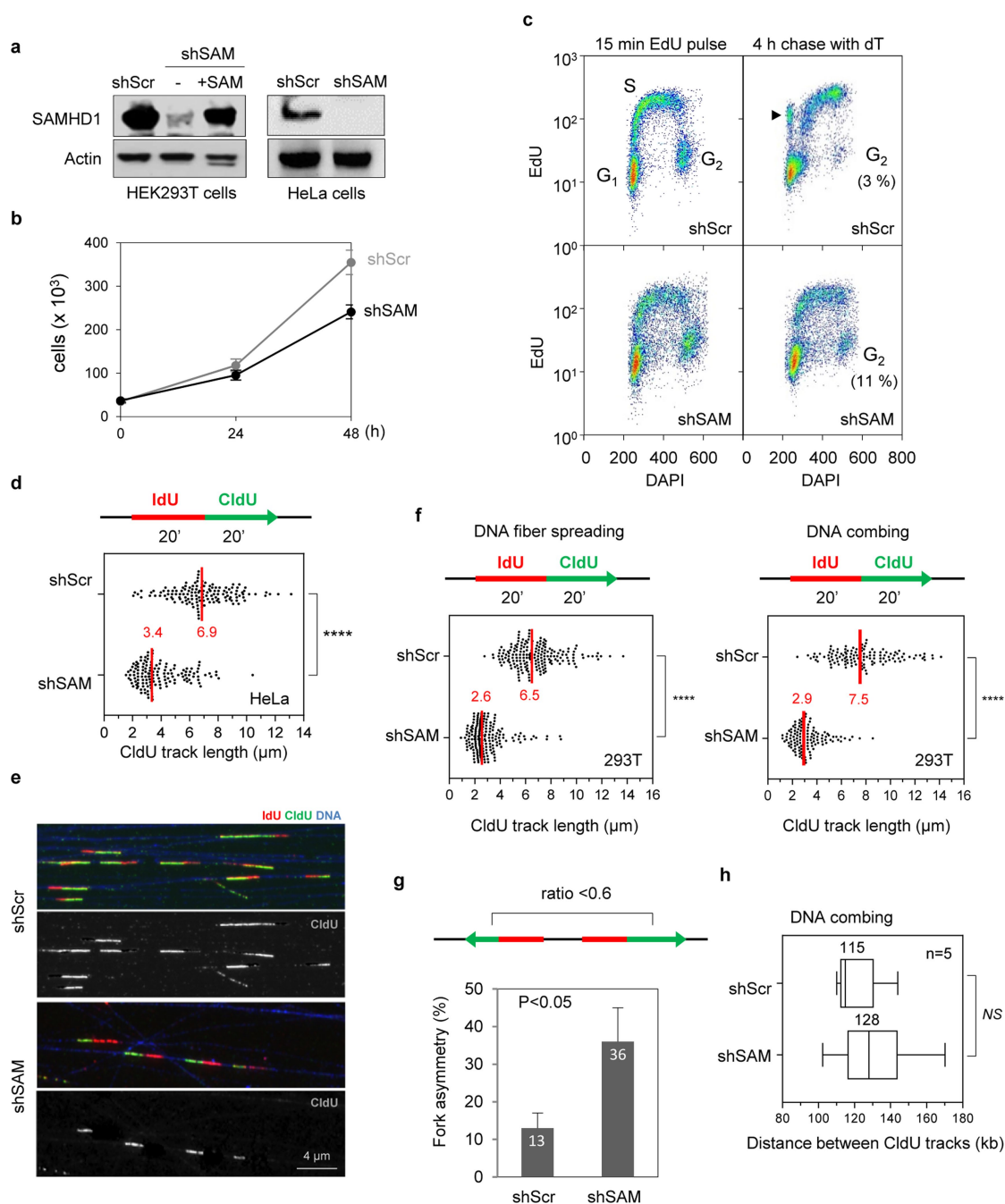
DNA in cytosolic DNA isolated from cells treated as in **g** was quantified by qPCR and normalized to *GAPDH*. Mean and s.d. from three independent experiments are shown. **i**, Levels of *STING* mRNA in cGAS-knockout, STING-knockout and control HeLa cells transfected with *SAMHD1* siRNA were measured by qRT–PCR 48 h after transfection. Mean and s.d. from three independent experiments are shown. **j**, Levels of cGAS protein in cGAS-knockout, STING-knockout and control HeLa cells transfected with *SAMHD1* siRNA were monitored by western blotting 48 h after transfection ($n = 3$). **k**, Levels of IRF3 protein in SAMHD1-depleted HeLa cells co-transfected with siRNAs against *STING* or *IRF3* ($n = 3$). **l**, HeLa cells were co-transfected for 48 h with siRNAs against *SAMHD1* and *STING* or *IRF3*. Levels of *ISG15* and *MX1* mRNA were analysed by qRT–PCR. Mean and s.d. from three independent experiments are shown. **m**, Expression levels of *STING* mRNA in HeLa cells co-transfected with siRNAs against *SAMHD1* and *STING* or *IRF3*. Mean and s.d. for three independent experiments are shown. **n**, HeLa cells were co-transfected for 48 h with siRNAs against *SAMHD1* and *STING* or *IRF3*. They were then treated with 4 mM HU for 8 or 18 h, washed with PBS and further cultured in fresh medium for 18 h. Expression of *ISG15* mRNA was quantified by qRT–PCR. Data are representative of three independent experiments. Mean and s.d. correspond to triplicates of a representative experiment. **o**, HEK293 cells (\pm *STING* siRNA) and CRISPR–Cas9-mediated STING-knockout THP-1 cells were transfected with *SAMHD1* siRNA. Expression of *MX1* mRNA was quantified by qRT–PCR. Mean and s.d. from three independent experiments are shown.



Extended Data Fig. 2 | See next page for caption.

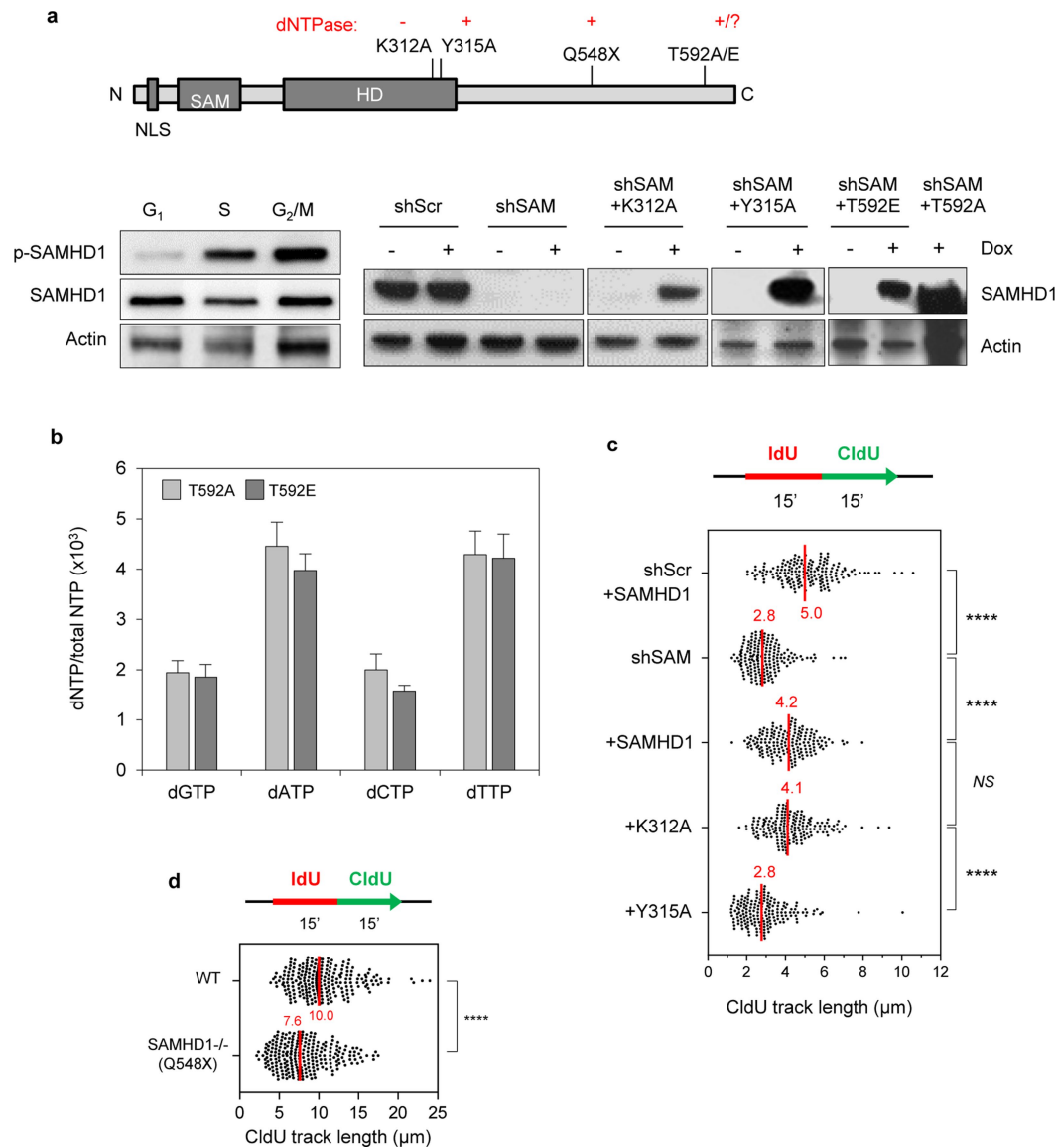
Extended Data Fig. 2 | SAMHD1 localizes to replication foci and binds nascent DNA. **a**, EdU (red) and SAMHD1 (green) foci in HeLa cells expressing HA-tagged SAMHD1 (SAMHD1-HA) or GFP-HA. Immunofluorescence microscopy was performed as indicated in Fig. 2a. Scale bar, 5 μm . $n = 2$. **b**, HeLa cells were labelled with 10 μM EdU for 10 min then processed for iPOND analysis. Proteins associated with nascent DNA were analysed by mass spectrometry. The number of peptides from SAMHD1 and other factors found associated with nascent DNA is indicated. The table summarizes the data from two independent experiments. **c**, Measurement of intracellular dNTP pools in shScr and shSAM HEK293T cells (two independent experiments). **d**, Production of recombinant His-xSamhd1 (see Supplementary Information) and characterization of the antibody raised against this protein. **e**, xSamhd1

associates with nascent DNA. *Xenopus* sperm DNA was incubated in *Xenopus* egg extract for 45 min then nascent DNA was labelled for 10 min with 40 μM biotin-16-dUTP in the absence or presence of 5 μM aphidicolin (Aph). Nascent chromatin was isolated on streptavidin beads and analysed by western blotting for the proteins indicated (see Methods). A representative experiment is shown ($n = 2$). **f**, Samhd1 binds chromatin in response to DSBs and aphidicolin. Left, *Xenopus* egg extracts were treated with the indicated doses of aphidicolin (μM), topotecan (T, 100 μM) or EcoRI ($\text{U } \mu\text{l}^{-1}$). Chromatin-bound proteins were then analysed by western blotting. Right, histograms show relative signal intensity of Samhd1 on chromatin. Data are mean and s.d. from four independent experiments.



Extended Data Fig. 3 | Fork progression is altered in the absence of SAMHD1. **a**, Western blot analysis of SAMHD1 protein in SAMHD1-depleted HeLa cells (using shSAM) and in SAMHD1-depleted HEK293T cells expressing an shRNA-resistant, full-length SAMHD1 (SAM) under the control of a doxycycline-inducible promoter. Expression of SAMHD1 was analysed 72 h after induction with doxycycline. **b**, Control (shScr) and SAMHD1-depleted (shSAM) HEK293T cells were seeded in 24-well plates. Cell number was determined by using trypan blue exclusion and haemocytometry. Mean and s.d. are shown from three independent experiments. **c**, SAMHD1 is required for normal S-phase progression. shScr and shSAM HEK293T cells were pulse-labelled with EdU for 20 min and chased with thymidine for 4 h before flow cytometry analysis. The arrowhead indicates the EdU-labelled cell population that completed DNA replication and came back to G₁ phase. The percentage of cells in G₂/M phase is indicated. $n = 2$. **d**, shScr and shSAM HeLa cells were labelled sequentially for 20 min with IdU and CldU and the length of CldU tracks ($n = 150$) was analysed by DNA fibre spreading. Median track lengths are indicated in red. **e**, Representative images of stretched DNA

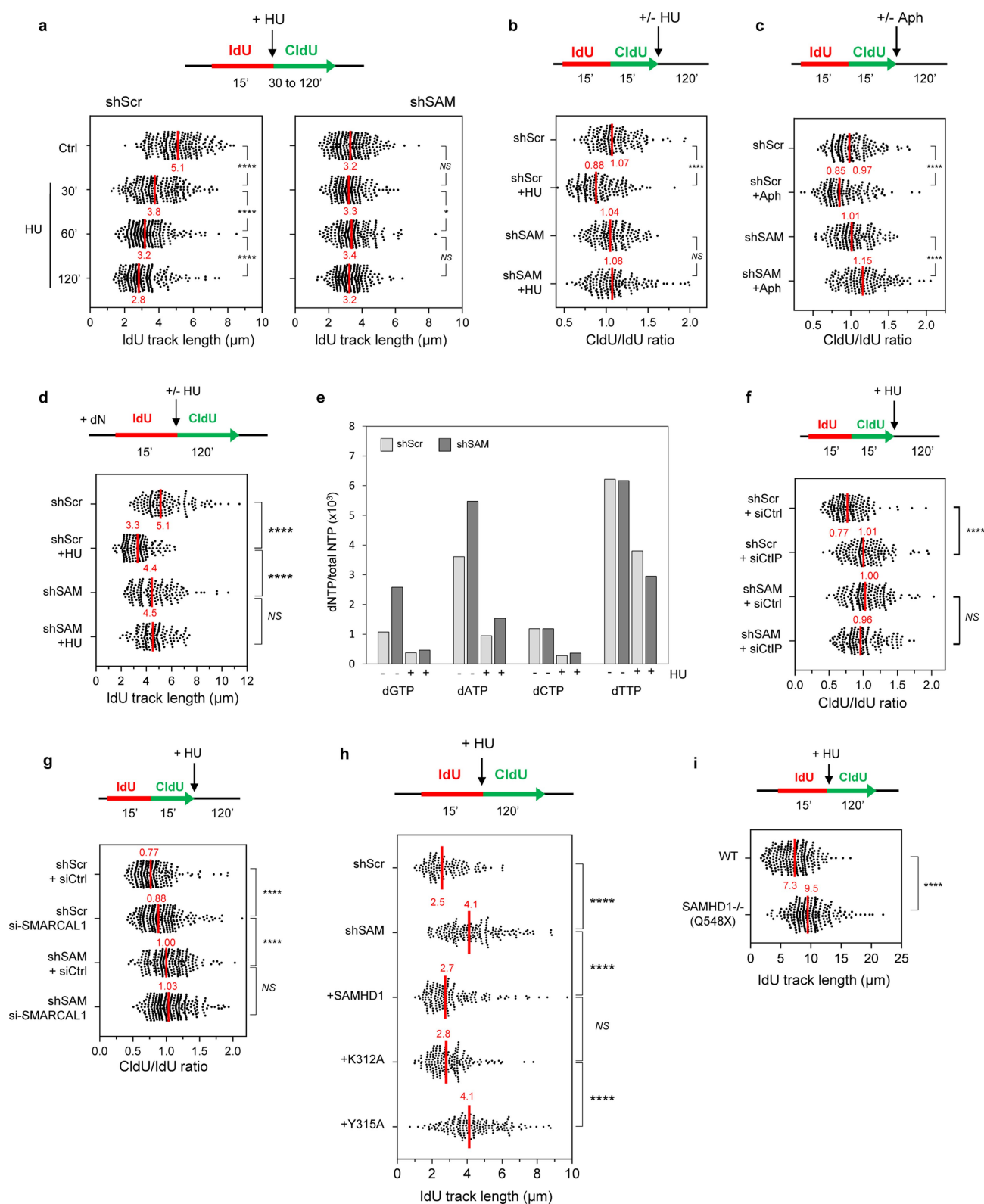
fibres. Red, IdU; green, CldU; blue, DNA. The green channel is shown separately for clarity. Scale bar, 4 μ m. **f**, DNA fibres from shScr and shSAM HEK293T cells sequentially labelled for 20 min with IdU and CldU were either stretched on glass slides (DNA fibre spreading; $n = 190$) or combed on silanised coverslips (DNA combing; $n = 130$) and then analysed by immunofluorescence microscopy. The length distribution of CldU tracks is shown. Median track lengths are indicated in red. **g**, DNA fibres from cells treated as in **f** were stretched by DNA combing and the distance between CldU tracks, which is indicative of the density of active origins, was determined for five independent experiments. Median distances are indicated. Whiskers correspond to 10–90 percentiles. **h**, SAMHD1-depleted cells show increased spontaneous fork arrest. The ratio of the shortest to the longest CldU track from cells treated as in **d** was calculated for each pair of divergent sister replication forks and the percentage of sister forks showing a ratio of less than 0.6 is shown ($n = 75$). Error bars indicate s.d. from three independent experiments. $P < 0.05$, Mann–Whitney rank sum test.



Extended Data Fig. 4 | SAMHD1 promotes fork progression

independently of dNTP pools. **a**, Map of SAMHD1 protein domains indicating the positions of the mutations analysed in this study. The level of phospho-SAMHD1 (T592) in HEK293T cells collected by FACS in G₁, S and G₂/M phases was determined with a phospho-specific antibody. Levels of wild-type and mutant SAMHD1 were also analysed by western blotting after induction of the genes with doxycycline (Dox) for 72 h. $n = 4$. **b**, SAMHD1-depleted HEK293T cells were complemented with the phosphomimetic (T592E) or non-phosphorylatable (T592A) mutants of SAMHD1. Intracellular dNTP pools were quantified as described in

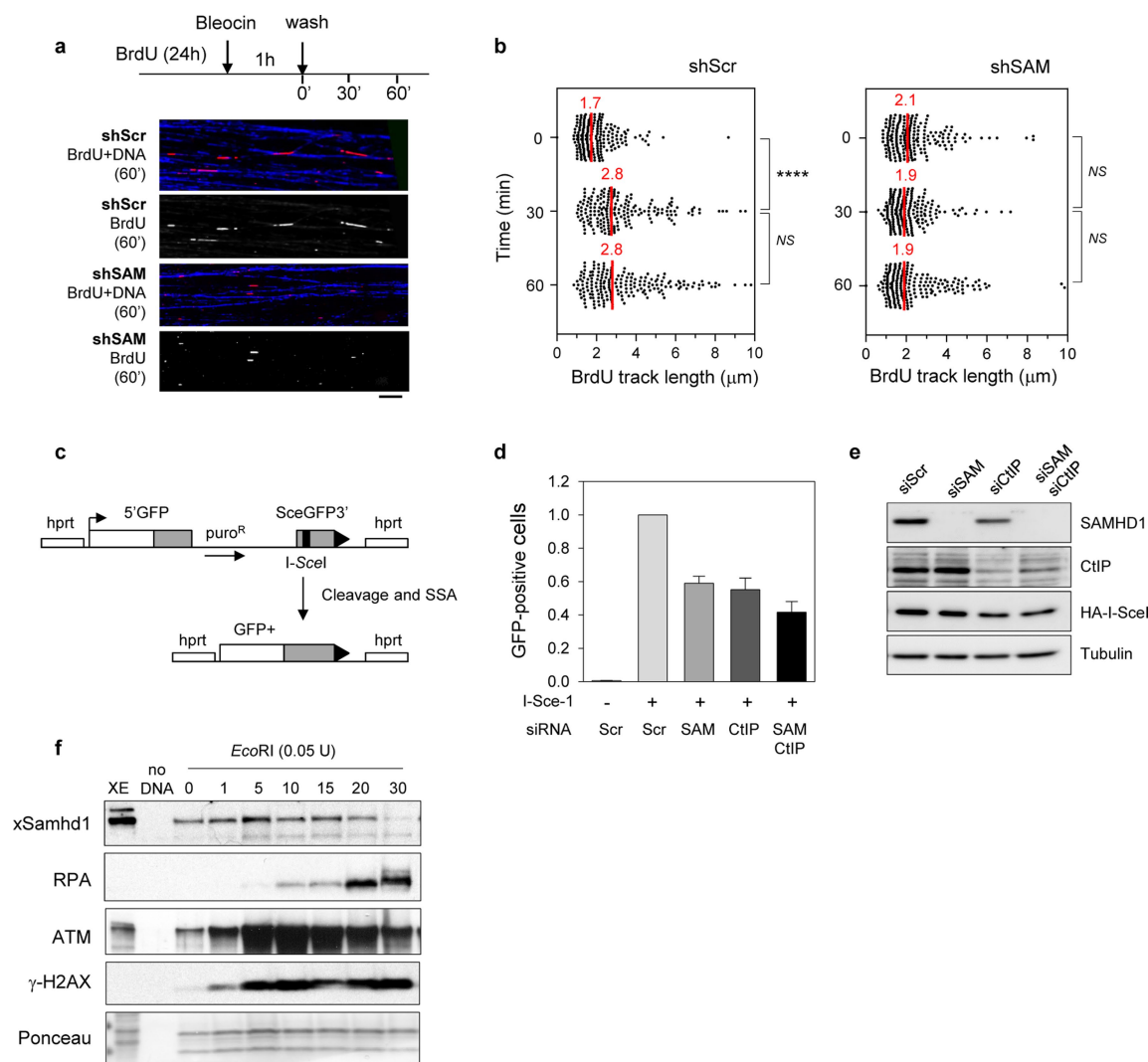
Methods. Median and s.d. are shown for five independent experiments. **c**, Wild-type SAMHD1 or the K312A and Y315A mutant forms were expressed in SAMHD1-depleted (shSAM) HEK293T cells and the cells were labelled sequentially with IdU and CldU for 15 min each. The lengths of the CldU tracks ($n = 180$) were measured on spread DNA fibres. **d**, Immortalized B cells from an *SAMHD1*^{-/-} patient with a homozygous Q548X mutation or a healthy donor (WT) were labelled with IdU and CldU and the lengths of CldU tracks ($n = 240$) were measured as in **c**. *** $P < 0.0001$, Mann-Whitney rank sum test.



Extended Data Fig. 5 | See next page for caption.

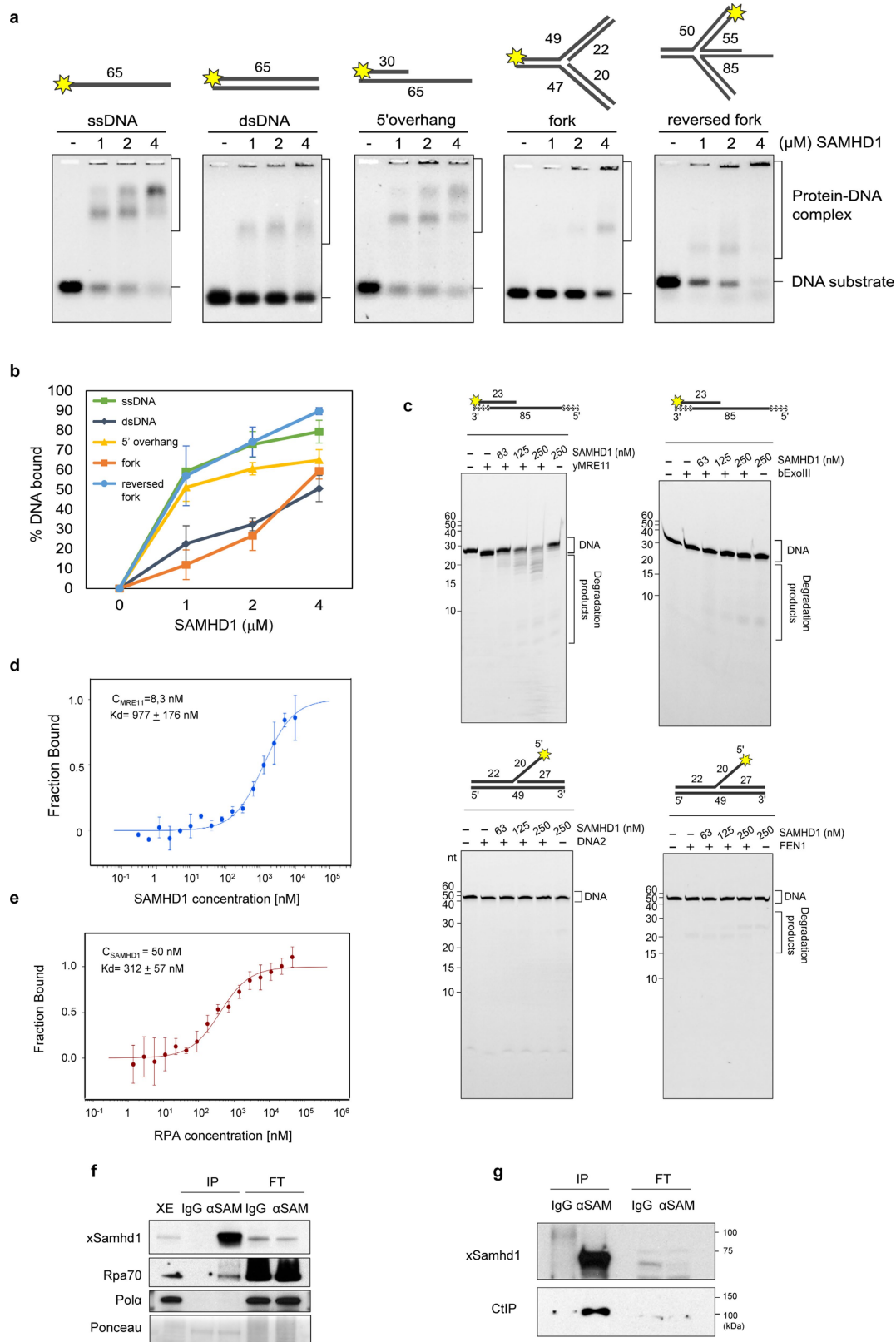
Extended Data Fig. 5 | Role of SAMHD1 in the degradation of nascent DNA at HU-arrested forks. **a**, Control (shScr) and shSAM HEK293T cells were labelled with IdU for 15 min and then exposed to 4 mM HU for 30, 60 or 120 min in the presence of CldU. The lengths of the IdU tracks ($n = 230$) were measured on spread DNA fibres. **b**, shScr and shSAM HEK293T cells were sequentially labelled for 15 min with IdU and for 15 min with CldU. Then, they were either collected immediately or treated for 2 h with 4 mM HU before DNA fibre analysis. The lengths of the IdU and CldU tracks ($n = 160$) were plotted as the ratio of CldU to IdU. **c**, shScr and shSAM HEK293T cells were treated as in **b**, except that HU was replaced with 1 μ M aphidicolin ($n = 160$). **d**, shScr and shSAM HEK293T cells were incubated for 120 min with a balanced mix of nucleosides (+dN) and then labelled with IdU and CldU in the presence of 4 mM HU, as indicated. The lengths of IdU tracks ($n = 160$) were measured on spread DNA fibres. **e**, shScr and shSAM HEK293T cells were treated for 2 h with 4 mM HU, or not treated, and intracellular dNTP pools were measured and expressed relative to intracellular rNTP pools in two independent experiments. **f**, shScr and shSAM HEK293T cells were transfected for 48 h with an siRNA against *CHP* (siCtIP) or a control (siCtrl) and then

labelled with IdU and CldU in the presence of 4 mM HU, as indicated. The lengths of the CldU and IdU tracks ($n = 160$) were measured on spread DNA fibres and plotted as the ratio of CldU to IdU track lengths. **g**, shScr and shSAM HEK293T cells were transfected with siRNA against *SMARCA1* or with a siCtrl for 48 h and then labelled with IdU and CldU in the presence of 4 mM HU, as indicated. The lengths of the CldU and IdU tracks ($n = 200$) were measured on spread DNA fibres and plotted as the ratio of CldU to IdU. **h**, Wild-type SAMHD1, the dNTPase-deficient mutant K312A or dNTPase-proficient mutant Y315A were expressed in shSAM HEK293T cells and the cells were labelled with IdU for 15 min and then with CldU in the presence of 4 mM HU for 2 h, as indicated. The lengths of the IdU tracks ($n = 140$) were measured on spread DNA fibres. **i**, Immortalized B cells from a *SAMHD1*^{-/-} patient with a homozygous Q548X mutation or a healthy donor (WT) were labelled with IdU and CldU in the presence of HU, as in **h**. The lengths of the IdU tracks ($n = 200$) were measured on spread DNA fibres. In **a–i**, median track lengths are indicated in red. * $P < 0.05$, **** $P < 0.0001$, Mann–Whitney rank sum test.



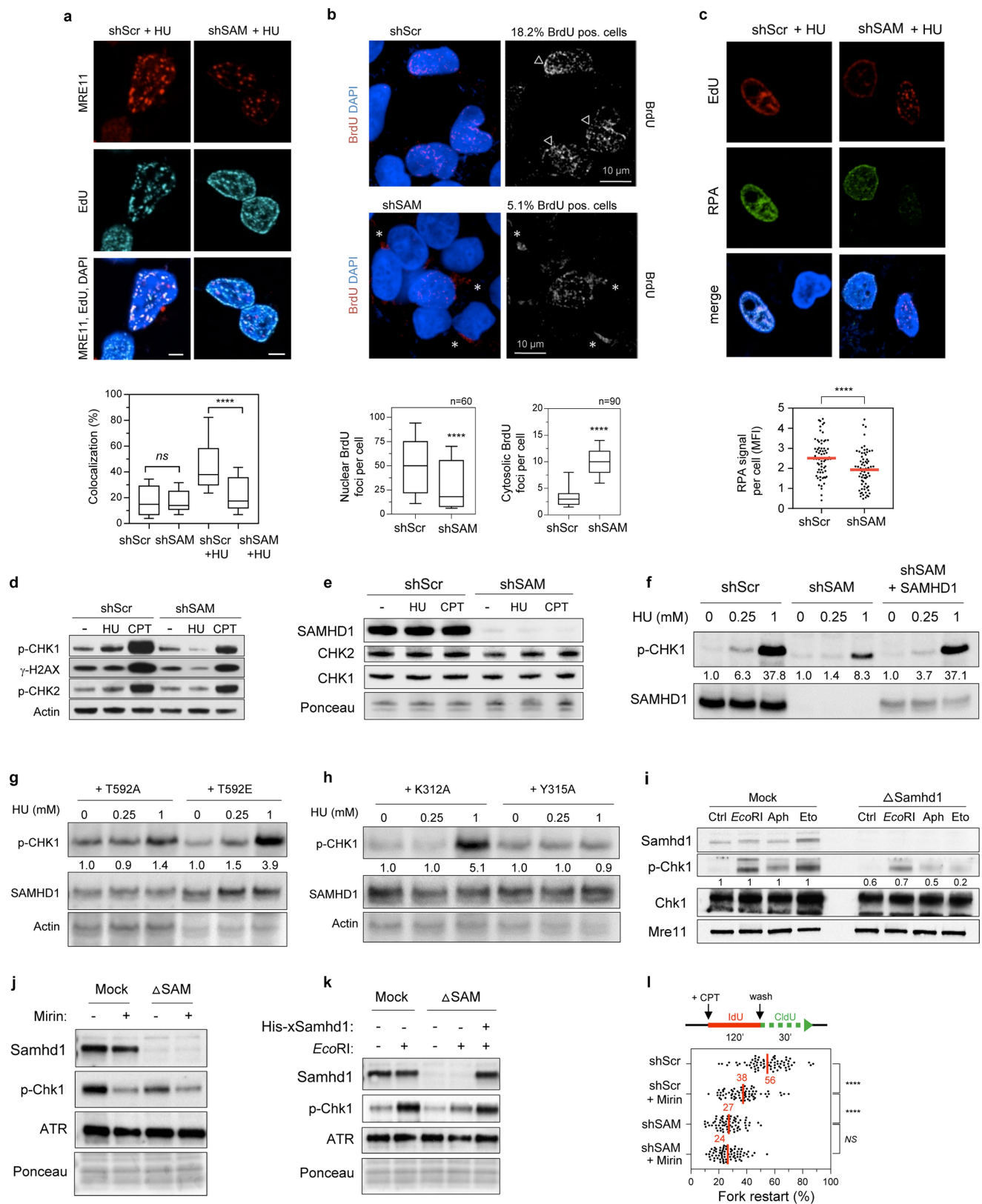
Extended Data Fig. 6 | SAMHD1 promotes the resection of DNA DSB ends. **a**, Analysis of DNA end resection at the level of individual DNA fibres with the SMART assay. Control (shScr) and SAMHD1-depleted (shSAM) HEK293T cells were grown for 24 h in the presence of BrdU to label genomic DNA and DSBs were induced with $5 \mu\text{g ml}^{-1}$ bleocin for 1 h. Cells were then washed and collected at the indicated times. DNA fibres were spread on glass slides and BrdU was detected without DNA denaturation. Representative images (three independent experiments) of BrdU tracks (red) 1 h after bleocin removal are shown. Scale bar, 5 μm. **b**, Quantification of BrdU track lengths ($n = 200$) in shScr and shSAM HEK293T cells treated with bleocin as in **a**. Median track lengths are indicated in red. **** $P < 0.0001$, Mann-Whitney rank sum test. **c**, Schematic of the U2OS single-strand annealing cell assay for DNA DSB repair. These cells carry a reporter vector in which an I-SceI site has

been incorporated into a GFP gene, such that single-strand annealing-mediated repair events result in GFP fluorescence. **d**, U2OS single-strand annealing cells were transfected with siRNAs against *SAMHD1*, *CtIP*, or both, or with a control scrambled siRNA (Scr). They were then transfected with a plasmid expressing HA-tagged I-SceI under the control of a CMV promoter. Percentages of GFP-positive cells were quantified by flow cytometry and were normalized to the control cells. Error bars denote s.d. of three independent experiments. **e**, Expression of SAMHD1, CtIP and HA-tagged I-SceI in the experiment shown in **d** were monitored by western blotting ($n = 3$). **f**, *Xenopus* sperm DNA was incubated in *Xenopus* egg extract in the presence of $0.05 \text{ U } \mu\text{l}^{-1}$ of EcoRI for the indicated times then the chromatin was purified and analysed by western blotting for the indicated proteins. A representative experiment is shown ($n = 3$).



Extended Data Fig. 7 | SAMHD1 binds MRE11 and stimulates its nuclease activity. **a**, Various fluorescently labelled DNA substrates (ssDNA, dsDNA, dsDNA with a 5' overhang, forked DNA and reversed fork DNA of the nucleotide lengths indicated) were incubated with a range of concentrations of purified SAMHD1 and the formation of protein-DNA complexes was determined by electrophoretic mobility shift assay (EMSA) as described in the Supplementary Information. Representative gel shift images are shown ($n = 3$). **b**, Quantification of the EMSA in **a**. **c**, SAMHD1 stimulates the nuclease activity of yeast MRE11, but not that of bacterial ExoIII and human FEN1 and DNA2. Error bars denote s.d. of three independent experiments. **d**, **e**, SAMHD1 binds MRE11

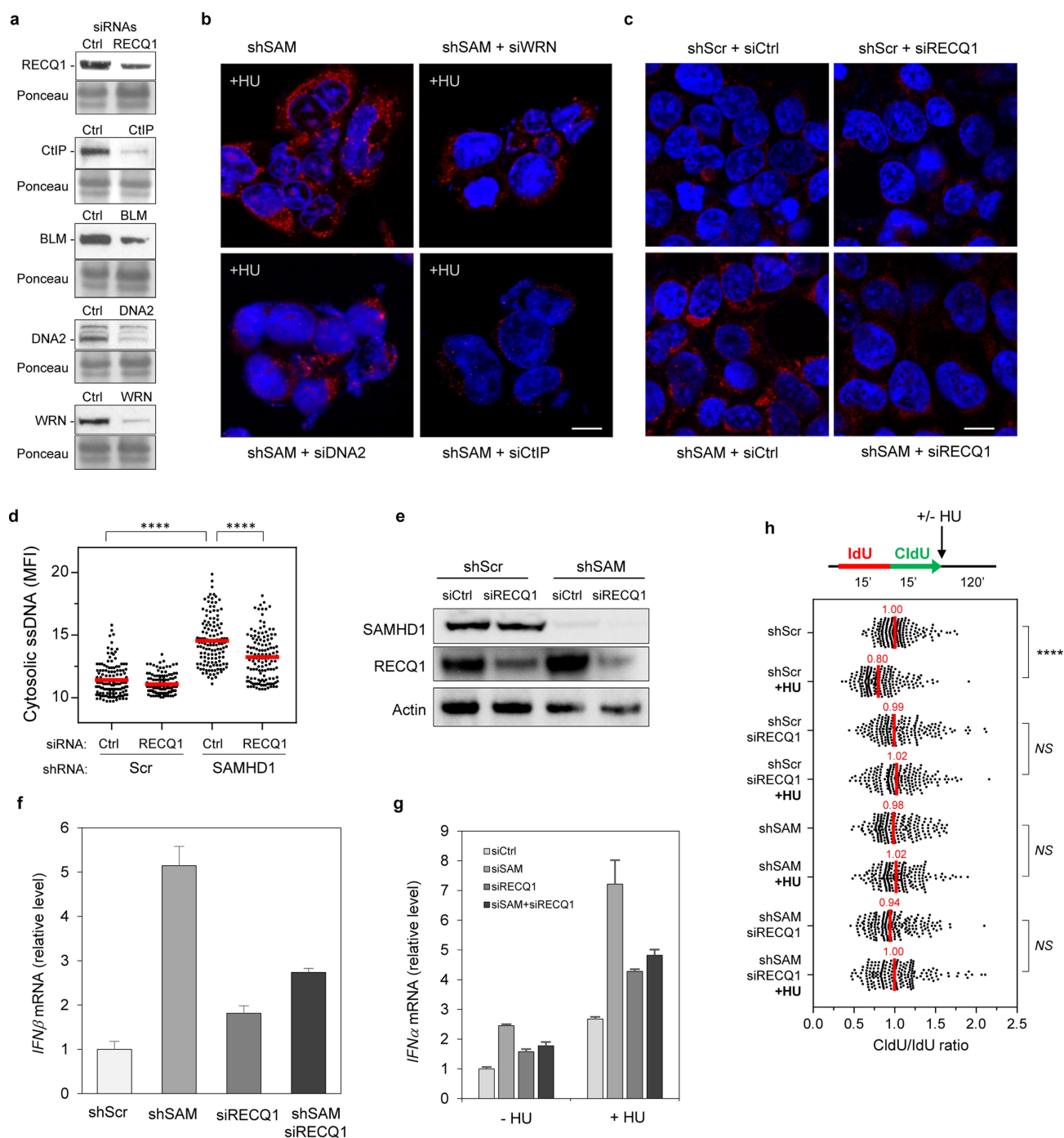
and RPA as monitored by microscale thermophoresis assay. Error bars denote s.d. of three independent experiments. **f**, Co-immunoprecipitation of Samhd1 and RPA from *Xenopus* egg extracts. Western blots are shown of the input egg extract (XE), the proteins immunoprecipitated (IP) by a control antibody (IgG), and an antibody against Samhd1 (α SAM), and the proteins that remained unbound (flow through, FT). A representative experiment is shown ($n = 2$). **g**, Co-immunoprecipitation of Samhd1 and CtIP from *Xenopus* egg extracts, as in **f** and described in the Supplementary Information. A representative experiment is shown ($n = 2$).



Extended Data Fig. 8 | See next page for caption.

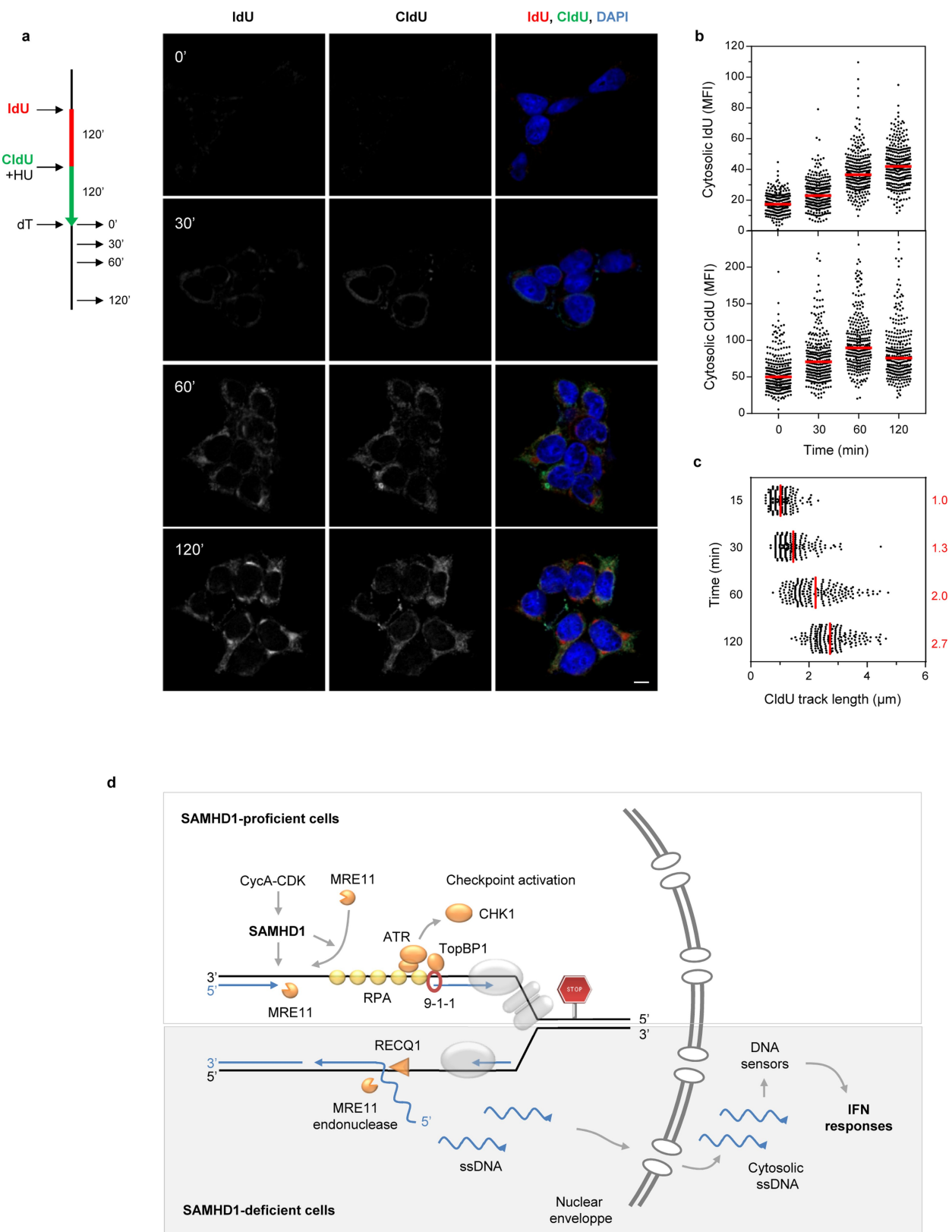
Extended Data Fig. 8 | SAMHD1 is required to recruit MRE11 and RPA to stalled forks and to activate the replication checkpoint. **a**, Top, control (shScr) and SAMHD1-depleted (shSAM) HEK293T cells were labelled with EdU for 10 min then grown without EdU for a further 2 h in the presence of 4 mM HU. Chromatin-bound MRE11 foci were detected by confocal microscopy and compared to EdU foci. Representative confocal images are shown. Scale bar, 5 μ m. Bottom, the co-localization of EdU and MRE11 foci was quantified by using the JACoP plugin of ImageJ ($n = 55$). Whiskers indicate the 10th and 90th percentiles. **** $P < 0.0001$, Mann–Whitney rank sum test. **b**, Top, foci (arrowheads) of ssDNA (red) in the nuclei (blue) of HU-treated shScr and shSAM HEK293T cells labelled for 24 h with 10 μ M BrdU. Representative images from two independent experiments are shown. Asterisks indicate BrdU-labelled cytosolic ssDNA. Bottom, quantification of nuclear and cytosolic BrdU foci per cell as in **a**. **c**, Top, shScr and shSAM HEK293T cells were incubated with 10 μ M EdU for 10 min and then treated for 2 h with 4 mM HU. RPA foci (green) were detected by using an anti-RPA1 antibody after CSK extraction. Bottom, mean fluorescence intensity of chromatin-bound RPA1 was quantified from 70 EdU-positive cells by using CellProfiler. **** $P < 0.0001$, Mann–Whitney rank sum test. Scale bar, 5 μ m. **d**, Immunoblots of CHK1 phosphorylated on S345 (p-CHK1), γ -H2AX phosphorylated on S139, and CHK2 phosphorylated on T68 (p-CHK2) in shScr and shSAM HEK293T cells treated for 60 min with 0.25 mM HU or 1 μ M CPT ($n = 3$). **e**, Expression of CHK1 and CHK2 proteins after HU or CPT treatment as indicated in **d**. **f**, shScr and shSAM HEK293T cells, and shSAM HEK293T cells expressing full-length SAMHD1 (shSAM + SAMHD1) were tested

for their ability to phosphorylate CHK1 on S345 upon exposure to 0.25 or 1 mM HU for 2 h. The fold of induction was normalized to untreated cells by quantifying the bands on the blots and calculating the ratios of p-CHK1 to SAMHD1 ($n = 3$). **g**, shSAM HEK293T cells expressing the non-phosphorylatable mutant T592A or the phosphomimetic mutant T592E of SAMHD1 were treated with 0.25 mM or 1 mM HU for 2 h. The amounts of p-CHK1 and SAMHD1 were analysed by immunoblotting and quantified as described in **f**; $n = 2$. **h**, shSAM HEK293T cells expressing the exonuclease positive (K312A) or negative (Y315A) mutants of SAMHD1 were treated with the indicated doses of HU for 2 h. The amounts of p-CHK1 and SAMHD1 were analysed by immunoblotting and quantified as described in **f**; $n = 2$. **i**, Depletion of Samhd1 from *Xenopus* egg extracts impairs the phosphorylation of Chk1 in response to EcoRI (0.05 U μ l⁻¹), aphidicolin (20 μ M) or etoposide (Eto, 30 μ M; $n = 3$). **j**, Samhd1 and Mre11 are required to activate Chk1 in Samhd1-depleted or mock-depleted *Xenopus* egg extracts treated with 0.05 U μ l⁻¹ of EcoRI. Mre11 was inhibited with 100 μ M mirin; this experiment was performed once. **k**, CHK1 activation in Samhd1-depleted *Xenopus* egg extracts upon formation of DSBs by addition of 0.05 U μ l⁻¹ EcoRI can be restored by the addition of approximately 25 nM of recombinant His-xSamhd1; a representative experiment is shown ($n = 2$). **l**, DNA fibre analysis of fork restart in shScr and shSAM HEK293T cells treated for 18 h with 10 μ M mirin and for 2 h with 1 μ M CPT. Fork restart (that is, formation of red and green tracks) was monitored 30 min after CPT removal ($n = 70$; three independent experiments). **** $P < 0.0001$, Mann–Whitney test.



Extended Data Fig. 9 | Depletion of RECQ1 prevents accumulation of cytosolic ssDNA in SAMHD1-depleted cells. a, RECQ1, CtIP, BLM, DNA2 and WRN proteins remaining 48 h after transfection with the corresponding siRNAs ($n = 2$). **b**, SAMHD1-depleted (shSAM) HEK293T cells were transfected with the indicated siRNAs for 48 h and then treated with 4 mM HU for 2 h. Cytosolic ssDNA was visualized by immunofluorescence microscopy. Representative confocal microscopy images are shown ($n = 2$). Scale bar, 5 μ m. **c**, Depletion of RecQ1 prevents the accumulation of cytosolic ssDNA in untreated shSAM cells. Cytosolic ssDNA was detected by immunofluorescence microscopy. Representative images are shown ($n = 3$). Scale bar, 5 μ m. **d**, Mean fluorescence intensity of cytosolic ssDNA in the cells **c** was quantified by using CellProfiler ($n = 130$). **** $P < 0.0001$, Mann-Whitney rank sum test. **e**, RECQ1 and SAMHD1 proteins remaining after siRNA transfection, as analysed by western blotting ($n = 3$). **f**, shScr and shSAM HEK293T

cells were transfected with siRNA against *RECQ1* (siRECQ1) for 48 h. Expression of *IFNβ* mRNA was quantified by qRT-PCR. Data shown are representative of three independent experiments. Error bars denote s.d. of triplicates. **g**, HeLa cells were transfected with siRNAs as indicated. They were cultured in the absence or presence of 4 mM HU for 8 h and then without HU for a total of 20 h before mRNA extraction. Expression of *IFNα* mRNA was analysed by qRT-PCR and normalized to *GADPH* mRNA. Data shown are representative of three independent experiments. Error bars denote s.d. of triplicates. **h**, shScr and shSAM HEK293T cells were transfected with siRNA against *RECQ1* for 48 h. The cells were sequentially labelled for 15 min with IdU and for 15 min with CldU. Then, they were either collected immediately or treated for 2 h with 4 mM HU before DNA fibre analysis. The lengths of the IdU and CldU tracks ($n = 200$) were plotted as the ratio of CldU to IdU. Median values are indicated in red. **** $P < 0.0001$, Mann-Whitney rank sum test.



Extended Data Fig. 10 | SAMHD1 depletion promotes accumulation of cytosolic nascent DNA under replication stress. a, SAMHD1-depleted HEK293T cells were labelled for 2 h with IdU and then for 2 h with CldU in the presence of 4 mM HU. They were then chased with thymidine for the indicated times and cytosolic IdU (red) and CldU (green) were visualized by confocal immunofluorescence microscopy. Representative images are shown. Scale bar, 5 μ m. **b**, Quantification of the cytosolic IdU and CldU signals in **a** by using CellProfiler ($n = 300$). **c**, DNA fibre analysis showing the increasing length of CldU tracks in HU-treated SAMHD1-depleted HEK293T cells. Median track lengths ($n = 190$) are

indicated in red. **d**, Model of the role of SAMHD1 at stalled replication forks. In SAMHD1-proficient cells (top), phosphorylation of SAMHD1 by the cyclin A (CycA)-CDK contributes to the MRE11-dependent resection of stalled replication forks and activates the ATR-CHK1 pathway at RPA-coated ssDNA, together with the DNA repair enzyme TopBP1 and the 9-1-1 (Rad9-Hus1-Rad1) complex. In SAMHD1-deficient cells (bottom), nascent DNA is displaced by the RECQ1 helicase and cleaved by an endonuclease, such as MRE11. The resulting ssDNA fragments accumulate in the cytosol and activate the type I IFN response.

Cryo-EM structure of the gasdermin A3 membrane pore

Jianbin Ruan^{1,2}, Shiyu Xia^{1,2}, Xing Liu^{1,3}, Judy Lieberman^{1,3} & Hao Wu^{1,2*}

Gasdermins mediate inflammatory cell death after cleavage by caspases or other, unknown enzymes. The cleaved N-terminal fragments bind to acidic membrane lipids to form pores, but the mechanism of pore formation remains unresolved. Here we present the cryo-electron microscopy structures of the 27-fold and 28-fold single-ring pores formed by the N-terminal fragment of mouse GSDMA3 (GSDMA3-NT) at 3.8 and 4.2 Å resolutions, and of a double-ring pore at 4.6 Å resolution. In the 27-fold pore, a 108-stranded anti-parallel β-barrel is formed by two β-hairpins from each subunit capped by a globular domain. We identify a positively charged helix that interacts with the acidic lipid cardiolipin. GSDMA3-NT undergoes radical conformational changes upon membrane insertion to form long, membrane-spanning β-strands. We also observe an unexpected additional symmetric ring of GSDMA3-NT subunits that does not insert into the membrane in the double-ring pore, which may represent a pre-pore state of GSDMA3-NT. These structures provide a basis that explains the activities of several mutant gasdermins, including defective mutants that are associated with cancer.

The gasdermin (GSDM) family, expressed in the skin, mucosa and immune antigen-presenting cells, triggers inflammatory programmed cell death (pyroptosis) and inflammatory cytokine secretion^{1–8}. There are six human GSDMs (GSDMA, GSDMB, GSDMC, GSDMD, GSDME (also known as DFNA5) and DFNB59 (also known as pejvakain)), and ten in mice, including three GSDMAs. GSDMs are cleaved by regulated processing that removes an inhibitory C-terminal fragment (GSDM-CT) to allow the N-terminal fragment (GSDM-NT) to bind to acidic lipids in the inner leaflet of mammalian cell membranes or on bacterial membranes to form pores. GSDMD is a substrate of inflammatory caspases^{1–8}, which are activated by inflammasomes that recognize invasive infection or intracellular danger signals^{9,10}, and GSDME is activated by caspase-3¹¹. The stimuli and proteases that activate the other GSDMs are largely unknown. To elucidate the mechanism of GSDM pore formation, we determined the cryo-electron microscopy (cryo-EM) structure of the mouse GSDMA3-NT pore.

Cryo-EM structure determination

We formed human GSDMD-NT and mouse GSDMA3-NT pores by cleaving the full-length proteins in the presence of phosphatidylserine-containing and cardiolipin-containing liposomes, respectively (Fig. 1a, Extended Data Fig. 1a, b). A detergent screen identified sodium cholate and C12E8 as suitable solubilizing agents for GSDMA3-NT and GSDMD-NT pores, respectively. Because GSDMA3-NT pores were more homogeneous in size and shape than GSDMD-NT pores (Extended Data Fig. 1c, d), we collected cryo-EM data from native GSDMA3-NT pores using a Talos Arctica microscope (Extended Data Fig. 1e), and from pores treated with HgCl₂ using a Titan Krios microscope (Extended Data Fig. 1f). HgCl₂ treatment was intended to label free Cys residues for validation of sequence assignment to the cryo-EM map.

For both datasets, top views of two-dimensional (2D) classified and averaged images showed mostly 27-fold symmetry (around 62% for native pores and around 70% for HgCl₂-treated pores), but there were also substantial numbers of pores with 26- (around 16% for native pores) and 28-fold (around 22% for native pores and around 30% for

HgCl₂-treated pores) symmetry, implying heterogeneity of oligomerization (Fig. 1b–e). For the native dataset, top-view classes of pores with 27-fold symmetry and all side views were used to perform three-dimensional (3D) classification without imposing any symmetry. One major class was further 3D-refined with 27-fold symmetry to a global resolution of 4.6 Å by gold-standard Fourier shell correlation (FSC) in Relion¹² (Fig. 1c, Extended Data Fig. 2a, Extended Data Table 1). For the HgCl₂-treated dataset, two rounds of 3D classification followed by 3D refinement led to maps at 3.8 Å and 4.2 Å resolution for the 27-fold- and 28-fold-symmetry pores, respectively (Fig. 1d, e, Extended Data Fig. 2b, c, Extended Data Table 1). Unexpectedly, the 4.6 Å cryo-EM map from native pores contains two rings with 27-fold-symmetry, with the top ring representing a membrane pore and the bottom ring with no membrane insertion (Fig. 1c). By contrast, the cryo-EM maps from HgCl₂-treated pores contain only the top ring (Fig. 1e). Local resolution distribution calculated in ResMap¹³ showed better resolution at the putative membrane-inserted region, which contained extended β-hairpins, in comparison to the juxtamembrane globular domain (Extended Data Fig. 2d, e).

Overall structure of the membrane pore

Analysis of the cryo-EM maps of the 27-fold membrane pore revealed a very large structure with an inner diameter of around 180 Å, an outer diameter of around 280 Å and a height of around 70 Å. We focus our discussion on the 3.8 Å-resolution pore, which had clearly defined secondary structures and large side chains (Fig. 2a, Extended Data Fig. 3, Supplementary Video 1). On the basis of the top views of 2D averages (Fig. 1b, d), the 26- and 28-fold pores would have similar dimensions that are only marginally smaller and larger, respectively. However, previously reported GSDMA3 pores assembled on cardiolipin-containing monolayer membranes appear to have 16-fold symmetry and a smaller interior width of 100–140 Å, despite the similar outer diameter of around 30 nm, as determined from 2D averages in negative-staining electron microscopy¹. Since native GSDMs assemble on lipid bilayers, which are present in our liposomes, rather than monolayers, and the determination of symmetry and dimensions from

¹Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. ³Department of Paediatrics, Harvard Medical School, Boston, MA, USA. *e-mail: wu@crystal.harvard.edu

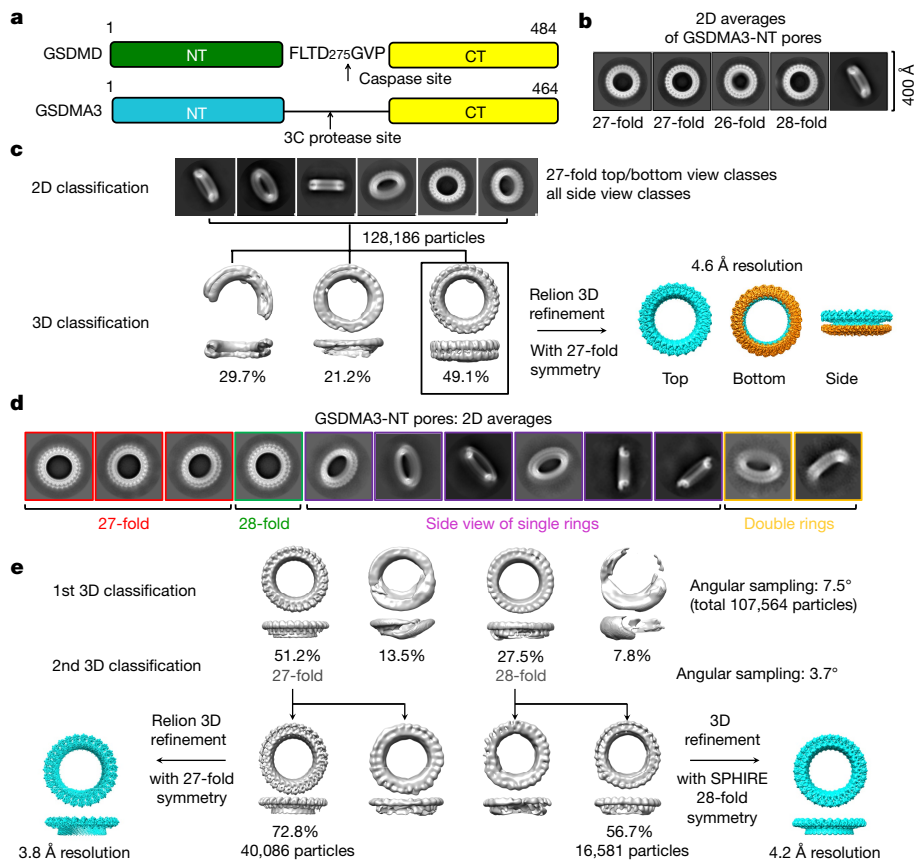


Fig. 1 | Structure determination of GSDMA3 pores. **a**, Cartoon diagram of human GSDMD and mouse GSDMA3 constructs used for in vitro pore reconstitution. **b**, 2D classification of GSDMA3 pores showing the presence of 26-, 27- and 28-fold symmetric pores. Each image shows an area of $400 \times 400 \text{ Å}^2$. **c**, 3D reconstruction of the 27-fold symmetric double-ring pore to 4.6 Å resolution. **d**, 2D classes of HgCl₂-treated GSDMA3 pores showing mostly single rings and a minor population of double rings. **e**, 3D reconstruction of 27- and 28-fold symmetric single-ring GSDMA3 pores to 3.8 and 4.2 Å, respectively.

low-resolution averages may be inaccurate, the structure we obtained here is more likely to resemble the architecture of the pore in the plasma membrane.

The predominant GSDMA3 pore with 27-fold symmetry is a complete anti-parallel β -barrel composed of 108 β -strands (Fig. 2a). Each subunit in the pore contributes four clearly separated long β -strands as two anti-parallel β -hairpins that align in a manner similar to the

digits of a left hand to form the transmembrane region (Fig. 2a, b). The β -strands that line the pore are from 15 to 22 residues long, and define a central β -barrel about 50 Å high, sufficient to traverse a lipid bilayer composed of lipid acyl chains with a thickness of about 30–40 Å. A previous analysis of transmembrane β -strands suggested that β -strands of between 6 and 22 residues in length can span a membrane¹⁴. The GSDMA3-NT globular region, or the ‘palm’ of the hand, shapes the rim

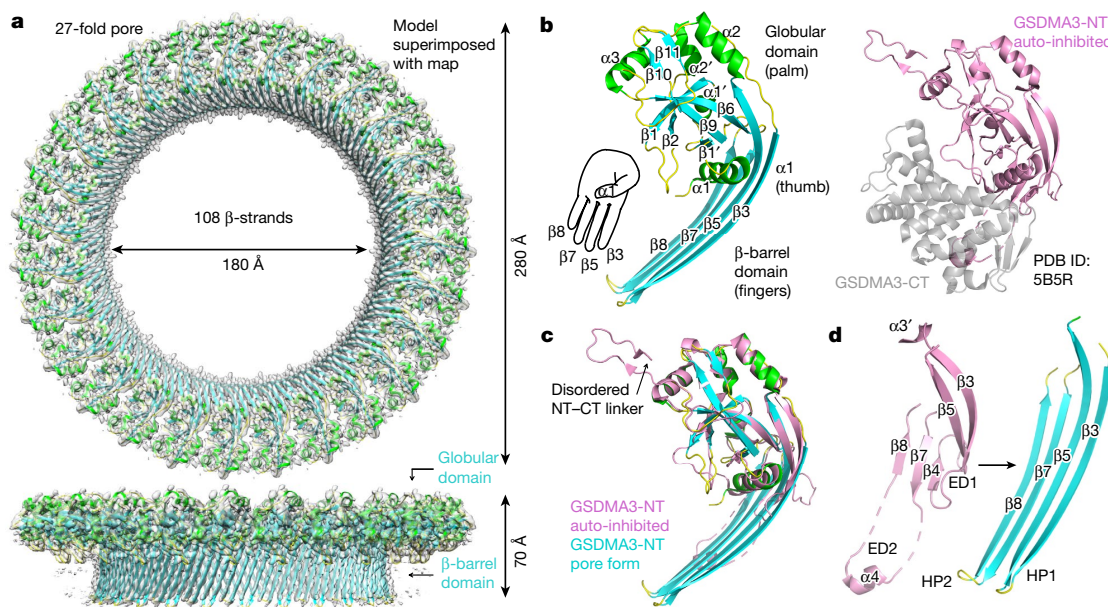


Fig. 2 | Structure and conformational changes of the GSDMA3 pore. **a**, Cryo-EM map (grey) superimposed onto the atomic model of the 27-fold symmetric GSDMA3 pore at 3.8 Å resolution. **b**, Ribbon diagram of GSDMA3-NT in the pore conformation (left) and crystal structure of

auto-inhibited GSDMA3¹ (right). **c**, Superposition of the auto-inhibited form and the pore form of GSDMA3-NT. **d**, Structural transitions that accompany the formation of the two β -hairpins HP1 and HP2 in the pore conformation. ED1 and ED2, extension domains 1 and 2, respectively.

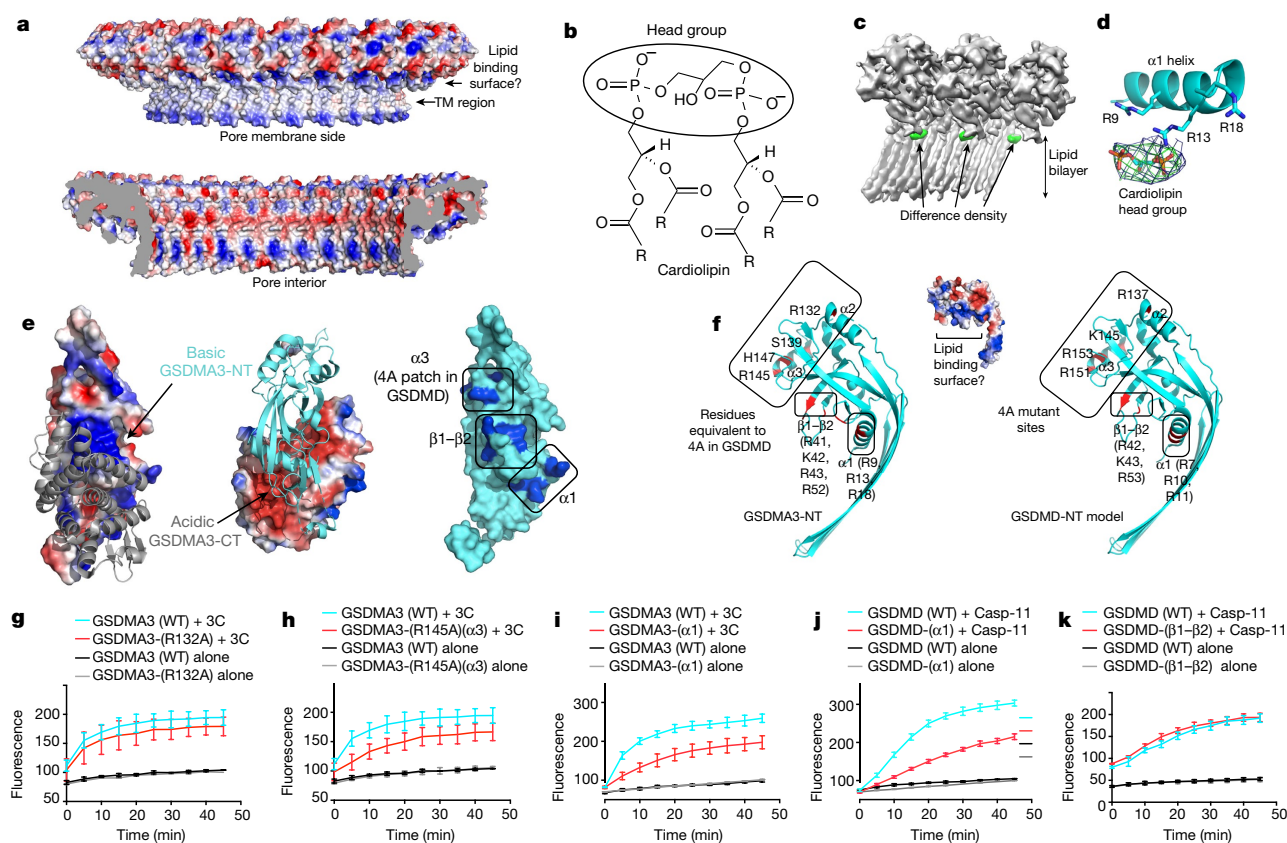


Fig. 3 | Mechanism of lipid recognition. **a**, Charge distribution in the GSDMA3 pore, viewed from the membrane side (top) and the pore interior (bottom). TM, transmembrane. **b**, Chemical structure of cardiolipin showing two phosphatidic acid moieties in the head group. The four R groups represent $C_{17}H_{33}$. **c**, Cryo-EM map of the GSDMA3 pore (grey) with the difference density map for potential bound lipid (green). **d**, Cardiolipin head group fitted into the lipid density in proximity to basic residues in the $\alpha 1$ helix. **e**, Potential lipid-binding patch. Left and middle, the auto-inhibited GSDMA3 structure¹, shown in electrostatic surface for the N-terminal domain and grey ribbon for the C-terminal domain

(left), and in cyan ribbon for the N-terminal domain and electrostatic surface for the C-terminal domain (middle). Right, the N-terminal domain in its auto-inhibited form shown as a surface diagram with the positively charged patches in blue. **f**, The basic patch in the pore form contains residues from helices $\alpha 1$ and $\alpha 3$ and the $\beta 1$ – $\beta 2$ region. Relevant residues are labelled for GSDMA3-NT (left) and for a model of human GSDMD-NT (right). **g–k**, Effects of GSDMA3 (g–i) and GSDMD (j, k) mutations on liposome-leakage activities, monitored by measuring dipicolinic acid (DPA)-chelating-induced fluorescence of released Tb^{3+} ion ($n = 3$ biological replicates). Error bars denote mean \pm s.d.

of the pore and contains both α - and β -elements, with the prominent $\alpha 1$ helix acting as the thumb (Fig. 2b).

Conformational transition upon membrane insertion

The conformation of the GSDMA3-NT pore exhibits a radical conformational change in comparison to its structure in the auto-inhibited, uncleaved protein (Protein Data Bank (PDB) ID: 5B5R)¹ (Fig. 2b–d). We named the secondary structures in the pore form on the basis of the auto-inhibited structure to facilitate comparison (Extended Data Fig. 4). Superposition of the two structures resulted in a root-mean-square deviation (r.m.s.d.) of 4.95 Å for 161 aligned C α atoms (Fig. 2c). Whereas the globular domain is largely unaltered, with a superimposed r.m.s.d. of 0.95 Å, marked conformational changes are associated with the formation of the two membrane-inserted β -hairpins. The entire $\beta 3$ – $\beta 4$ – $\beta 5$ region extends into the first tight transmembrane β -hairpin, and the $\alpha 3'$ and the disordered region preceding it in the auto-inhibited conformation are pulled into the new $\beta 3$ strand (Fig. 2c, d, Supplementary Video 2). The entire $\beta 7$ – $\alpha 4$ – $\beta 8$ region stretches out into the second transmembrane β -hairpin, including the associated loops and disordered segments in the auto-inhibited structure (Fig. 2c, d, Supplementary Video 2). The $\alpha 4$ helix that binds to the C-terminal fragment (GSDMA3-CT) in the uncleaved structure (Fig. 2d) forms part of the tip of the second β -hairpin. Because of the structural transformation, we named the $\beta 4$ and $\alpha 3'$ region ‘extension domain 1’ (ED1), the $\alpha 4$ region ‘extension domain 2’ (ED2), and the newly formed long β -strands ‘hairpin 1’ (HP1) and ‘hairpin 2’ (HP2) (Fig. 2d).

Acidic lipid binding by the $\alpha 1$ helix

Analysis of the electrostatic surface of the GSDMA3-NT pore revealed that in the transmembrane region, the side of the pore that faces the membrane is largely hydrophobic, as expected, whereas hydrophilic residues form positively and negatively charged stripes on the inner side of the pore conduit (Fig. 3a). A positively charged surface patch also forms adjacent to the proposed transmembrane region, in the globular domain of the pore (Fig. 3a). When we used a difference density map to locate additional densities, we found a strong density at 6.6σ adjacent to the $\alpha 1$ helix, which fits well to the double phosphate head group of cardiolipin (Fig. 3b–d), the acidic lipid we used to reconstitute the pore on liposomes. We hypothesize that the modelled cardiolipin head group interacts with the basic residues on helix $\alpha 1$ (Fig. 3d).

To confirm the role of helix $\alpha 1$ in lipid interaction, we inspected the membrane-proximal, positively charged surface of the GSDMA3 pore, which is made up of basic residues from helix $\alpha 1$, helix $\alpha 3$ and strands $\beta 1$ and $\beta 2$ of the pore form of GSDMA3-NT (Fig. 3e, Extended Data Fig. 4). These regions are masked by GSDMA3-CT in the auto-inhibited full-length structure (Fig. 3e), consistent with binding of GSDMD-NT, but not full-length GSDMA3 or GSDMA3-CT, to acidic lipids^{1–4,8}. Mutation to Ala of a cluster of four conserved basic residues on helices $\alpha 2$ and $\alpha 3$ in mouse GSDMD (GSDMD(4A)) that correspond to R137, K145, R151 and R153 of human GSDMD (Fig. 3f, right) compromised lipid binding, liposome leakage and pyroptosis². When these mutations were combined, the loss of both R151 and R153 on $\alpha 3$ was the most defective combination, whereas the combined loss of R137

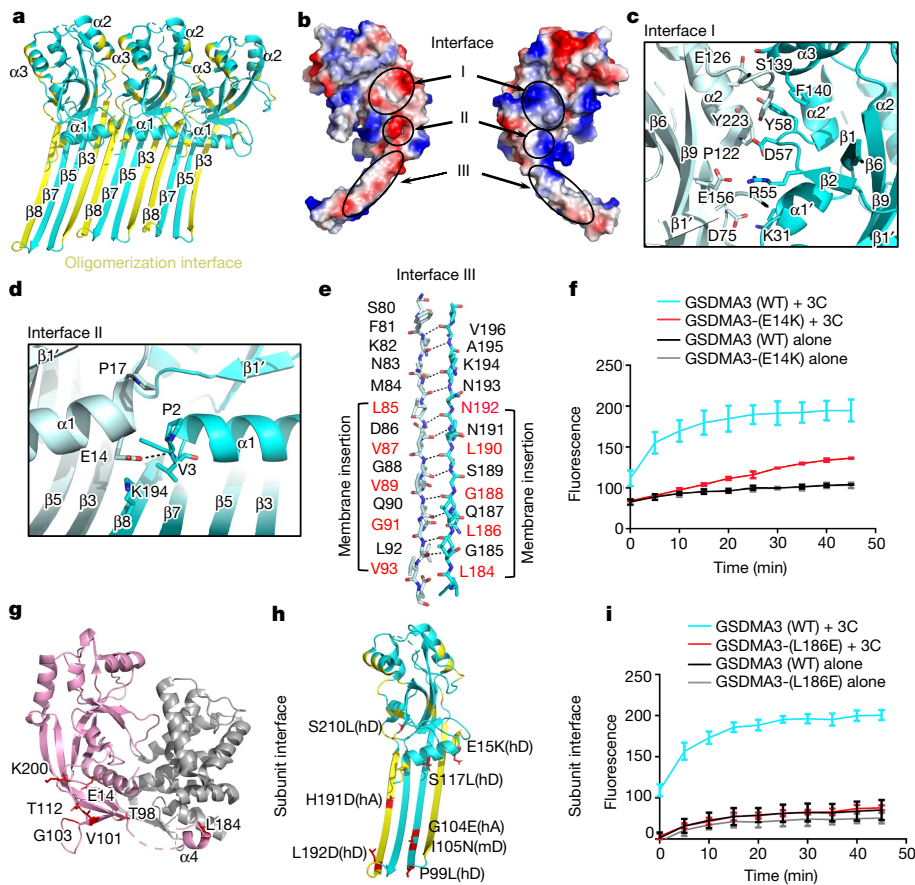


Fig. 4 | Mechanism of oligomerization and membrane insertion.

a, Three contiguous subunits in the GSDMA3 pore. Residues that participate in oligomerization are labelled in yellow. **b**, Electrostatic potential map of a GSDMA3-NT subunit in two side views showing the oligomerization interface with three major patches labelled.

c–e, Main residues on interface I (**c**), interface II (**d**) and interface III (**e**). **f**, Oligomerization mutant E14K exhibits compromised pore formation compared to wild-type (WT) GSDMA3 ($n = 3$ biological replicates). Error bars denote mean \pm s.d. **g, h**, The locations of previously reported and cancer-associated mutations on both auto-inhibited (**g**) and pore (**h**) conformations on GSDMA3. hA, human GSDMA; hD, human GSDMD; mD, mouse GSDMD. **i**, Mutation of L186 in the membrane insertion region abolishes pore formation and liposome leakage ($n = 3$ biological replicates). Error bars denote mean \pm s.d.

and K145 only minimally impaired cell death². Two of the four basic residues in GSDMD(4A) are conserved in GSDMA3 as R132 in $\alpha 2$ and R145 in $\alpha 3$ (Fig. 3f, left). To test the importance of these residues, we generated R132A and R145A mutants of GSDMA3. Liposome leakage assays showed that the R132A mutation on $\alpha 2$ barely compromised pore formation (Fig. 3g), whereas the R145A mutation on $\alpha 3$ or the R132A/R145A double mutation weakened liposome disruption (Fig. 3h, Extended Data Fig. 5a), consistent with the defective phenotype of corresponding mutations on GSDMD².

To test the effect of the positively charged juxtamembrane patches formed from residues in helix $\alpha 1$ and the $\beta 1$ – $\beta 2$ region, we generated a triple mutant of GSDMA3 containing R9A, R13A and R18A (GSDMA3($\alpha 1$)), and a quadruple mutant containing R41A, K42K, R43A and R52A (GSDMA3($\beta 1$ – $\beta 2$)) (Fig. 3f, left). The triple mutant was expressed and could be cleaved by the 3C protease with similar kinetics to those of the wild-type protein (Extended Data Fig. 5b), suggesting that the mutation did not radically alter the structure of the full-length protein, whereas the quadruple mutant could not be evaluated because it was not expressed. GSDMA3- $\alpha 1$ cleaved by 3C protease had reduced membranolytic activity in the liposome leakage assay (Fig. 3i). Corresponding Ala mutations were also introduced into human GSDMD at R7, R10 and R11 (GSDMD($\alpha 1$)), and at R42, K43 and R53 (GSDMD($\beta 1$ – $\beta 2$)) (Fig. 3f, right). Both mutants were expressed and could be purified. Liposome leakage assays showed that similar to GSDMA3($\alpha 1$), the GSDMD($\alpha 1$) mutant also showed compromised pore formation in comparison to the wild-type protein (Fig. 3j). The importance of the $\alpha 1$ helix is also supported by the lack of cell death when part of $\alpha 1$ is deleted⁴. By contrast, the GSDMD($\beta 1$ – $\beta 2$) mutation did not decrease liposome leakage (Fig. 3k).

Thus, these mutational studies suggest that the $\alpha 1$ and $\alpha 3$ juxta-membrane basic patches, but not basic residues in the $\beta 1$ – $\beta 2$ region, are potential candidates for lipid binding. Because $\alpha 3$ also participates in oligomerization between subunits (see below), and is further away from the membrane than $\alpha 1$, we argue that the $\alpha 1$ helix is likely to be

the major lipid-binding site for GSDM pore formation. The difference density near $\alpha 1$ may represent the acidic head group of cardiolipin, which may interact strongly with GSDMA3-NT and stay associated with the pore after detergent extraction. We do not know whether the acidic lipid interaction is only important for inducing conformational changes in GSDM or whether it is also essential for stabilizing the membrane pore.

Oligomerization interface

Analysis of the oligomerization interface of the GSDMA3-NT pore revealed extensive interactions of both the inserted β -strands and the associated globular domains (Fig. 4a). The β -strands have a tilt angle of about 20° relative to the pore height direction. If one draws the β -strand hydrogen bonding pattern across the β -barrel, there is a two-residue shift from one subunit to the next subunit (Extended Data Fig. 6), giving rise to a total displacement of residues, or a shear number¹⁵, of $27 \times 2 = 54$ for the entire 27-fold symmetric pore. Each subunit interface buries a total of $1,600 \text{ \AA}^2$ surface area, with key interaction areas dividable into three patches, I, II and III (Fig. 4b, Extended Data Fig. 4). In the first patch, residues between the neighbouring globular domains form both hydrophobic and charged interactions (Fig. 4c). The interaction contains mainly residues from helix $\alpha 3$ of one subunit and the region around $\alpha 2$ and $\beta 11$ of its neighbouring subunit. In the second patch, the $\alpha 1$ helix from one subunit juxtaposes end-on with the $\alpha 1$ helix from the next subunit through hydrogen-bonding and hydrophobic interactions (Fig. 4d). The third major subunit interface runs along the neighbouring $\beta 3$ and $\beta 8$ strands between the subunits (Fig. 4e). Notably, in the inserted β -strands, residues facing the membrane are hydrophobic, while those facing the pore are mostly hydrophilic or charged (Fig. 4e).

Previous studies showed that E15K and L192D mutations of GSDMD-NT, and the corresponding mutations E14K and L184D of GSDMA3-NT (Extended Data Fig. 4), strongly reduced pyroptosis¹. Consistent with these results, we found that the E14K mutant of

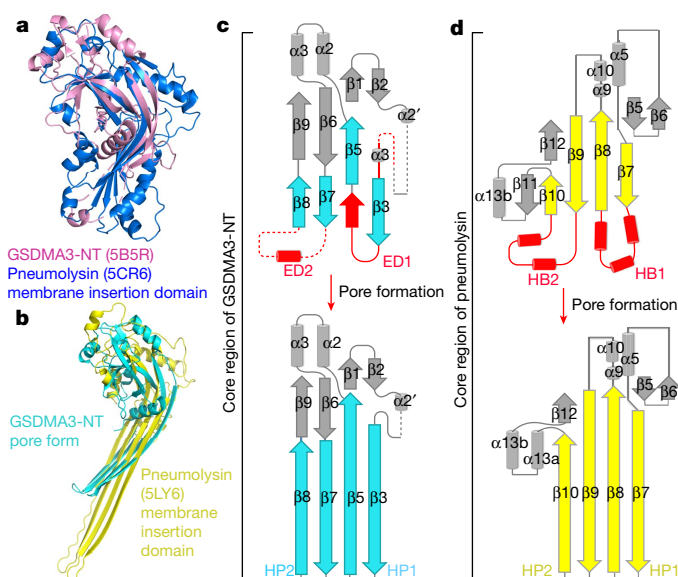


Fig. 5 | Comparison with MACPF/CDC proteins. **a, b**, Superposition of GSDMA3-NT with the membrane insertion domain of pneumolysin in the soluble form (**a**) and in the pore conformation (**b**), showing the alignment of the central β -sheets. The C-terminal tail of GSDMA3-NT that links to GSDMA3-CT and part of the pneumolysin structure are omitted for clarity. **c, d**, Different structural topologies of GSDMA3-NT (**c**) and pneumolysin (**d**) in both soluble and pore forms. ED, extension domain; HB, helix bundle; HP, β -hairpin. Regions that are straightened by the insertion are highlighted in red.

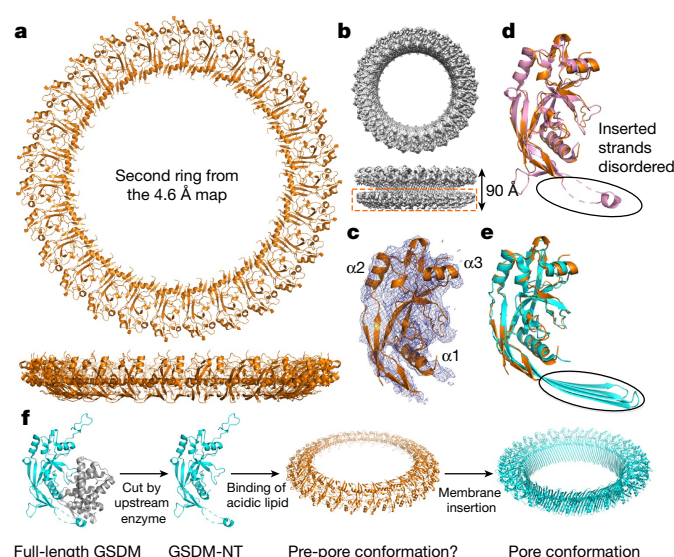


Fig. 6 | A possible GSDMA3 pre-pore conformation. **a, b**, Model of the additional ring (**a**) and its cryo-EM density in the double-ring GSDMA3 pore architecture (**b**). **c**, Cryo-EM density of an individual subunit superimposed on the model. **d, e**, Superposition of the subunit structure in the additional ring with that of GSDMA3-NT in the auto-inhibited conformation (pink)¹ (**d**) and the pore conformation (cyan) (**e**). **f**, A hypothetical model for GSDM pore formation. Upon cleavage by an activating enzyme, GSDM-NT monomers may first bind to membrane lipids and then oligomerize to form a soluble pre-pore before membrane insertion.

GSDMA3 compromised liposome leakage in vitro (Fig. 4f). The inhibitory effects of E14 and L184 mutations on GSDMA3 cannot be explained by its full-length structure because E14 is close to the N-terminal domain–C-terminal domain interface and L184 interacts closely with the C-terminal domain (Fig. 4g), and mutations in the C-terminal domain near E14 or L184 constitutively activate GSDMA3 and GSDMD^{1,6}. However, in the GSDMA3-NT pore structure, E14 and L184 of GSDMA3 are involved in oligomerization at the globular domain and the β -barrel domain, respectively (Fig. 4d, e, g, h); this explains how these mutations compromise pore formation, even though they reduce auto-inhibition. L184 is also involved in membrane insertion and its mutation to an acidic residue could therefore also directly affect this process. To further test the structural model, we mutated another oligomerization or insertion residue L186 to obtain the L186E mutant. As we predicted, the mutation almost completely eliminated the liposome leakage activity of proteolytically cleaved GSDMA3 (Fig. 4i).

A forward genetic screen with randomly mutated mice identified an I105N mutation of mouse GSDMD as defective in the intracellular LPS response in vivo, and in membrane permeabilization in vitro^{7,8}. The I105N mutation in mouse GSDMD corresponds to I104N in human GSDMD and V101N in GSDMA3 (Extended Data Fig. 4). V101 localizes at $\beta 5$ of the pore form and is involved in membrane insertion (Fig. 4h), explaining the impaired function in the mutant GSDMD. The OASIS cancer genome site (<http://www.oasis-genomics.org/>) lists a number of point mutations in human GSDMA and GSDMD from patient samples, including mutations in GSDMA that correspond to G103E and N192D in GSDMA3, and GSDMD mutations E15K, E15Q, P99L, S117L and S210L that correspond to mutations on residues E14, T98, T112 and K200 in GSDMA3 (Fig. 4d, g, h). E14 of GSDMA3 is at the oligomerization interface of the pore globular domains, whereas the remaining residues are all involved in membrane insertion and/or oligomerization at the β -barrel domain (Fig. 4h), suggesting a potential mechanism of tumour potentiation through reduced pyroptotic activity of GSDMs.

Comparison with cytolysins

The large β -barrel architecture of the GSDMA3 pore described here shares mechanistic similarities with pores formed by bacterial cytolysins, complement proteins and the pore-forming protein perforin from cytotoxic T lymphocytes, which belong to the ‘membrane attack complex perforin-like’/‘cholesterol dependent cytolysin’ (MACPF/CDC) family^{16–18}. Whereas many structures in the family have been determined in their soluble forms, the only high-resolution structure of the pore form is that of pneumolysin^{19,20}. A DALI search²¹ using the GSDMA3-NT domain structures in both the full-length context and the pore form identified weak structural homology at the central β -sheets to these proteins, including the soluble form and the pore form of pneumolysin^{19,20} (Fig. 5a, b), with the highest structural homology to perfringolysin O, with a r.m.s.d. of 7.3 Å for 134 aligned residues.

Despite the structural similarity at the central β -sheets, we hypothesize that the GSDM and MACPF/CDC families either evolved independently or are so distantly related that divergent evolution at the sequence level and unexpected conserved structural features are not detected. We present three major differences that support this idea. First, previous analysis showed that the MACPF and CDC family proteins contain three conserved glycine residues²², which are absent in GSDMs (Extended Data Fig. 7). Second, the structural homology between MACPF and CDC spreads from the central β -sheets to the clusters of α -helices that switch to the membrane-spanning β -conformation, yielding much lower r.m.s.d. for many more aligned residues than with GSDMA3²². Third, the membrane insertion mechanism is different. For GSDMA3, one β -hairpin is formed from straightening the $\beta 7$ and $\beta 8$ strands into the connected $\alpha 4$ and disordered loops, while the other hairpin comes from both the elements in between $\beta 3$ and $\beta 5$, and elements preceding $\beta 3$ that include the $\alpha 3'$ helix and a disordered loop in the auto-inhibited conformation (Fig. 5c). By contrast, in pneumolysin, straightening into the connected cluster of helices from the central β -sheet alone is responsible for the formation of the long β -hairpins for membrane insertion and β -barrel formation¹⁹ (Fig. 5d, Supplementary Video 3).

The soluble pore may represent a pre-pore

The native cryo-EM map of GSDMA3-NT shows the existence of an additional pore, creating a yo-yo-shaped double-ring pore (Figs. 1c, 6a, b). Unlike the membrane pore, the bottom ring does not insert into the membrane, and we therefore call it the soluble pore. We do not know the physiological significance of the double-ring structure. On the other hand, most HgCl₂-treated GSDMA3 pores contain only the membrane-inserted ring (Fig. 1e), suggesting that the double-ring structure is sensitive to the biochemical condition, and could be an artefact of in vitro reconstitution.

Although the cryo-EM density for the additional ring is much poorer, we could fit the N-terminal part of the full-length auto-inhibited GSDMA3 structure¹ rigidly into the density (Fig. 6c). Consistent with lack of membrane insertion, most of the β -strands in the transmembrane region of the pore form, as well as the N-terminal–C-terminal linker that harbours the 3C cleavage site, are disordered in this fitted GSDMA3-NT structure (Fig. 6d, e). Although detailed analysis of the GSDMA3-NT conformation and its subunit association in the soluble pore would require improved resolution of this region, the existence of soluble pre-pore conformations for some MACPF/CDC family members¹⁹ suggests that the soluble GSDMA3-NT pore could represent a pre-pore state. Notably, a mutant of aerolysin forms asymmetric pores with a stacked quasi-pore and a soluble pore^{23,24} that resemble our double-ring native GSDMA3 pore structure.

Summary

The cryo-EM structure of the GSDMA3-NT pore revealed that the N-terminal fragment of GSDM undergoes marked conformational changes to form the pore structure upon binding to acidic inner leaflet or bacterial lipids. Difference density and structure-based mutagenesis suggested that helix α 1 is the major site of acidic lipid interaction, which is required for pore formation. We do not know the function of the second ring or whether it is part of the physiological pore in cells or an artefact of in vitro reconstitution; however, we speculate that it may represent a pre-pore structure that precedes membrane insertion (Fig. 6f).

In some circumstances, membrane damage caused by bacterial cytolytins, complement proteins or perforin can be repaired by the ubiquitous plasma membrane repair pathway, which is triggered by a calcium influx from the extracellular milieu into the damaged cell^{25,26}. In the case of perforin, repair enables target cells to die by apoptosis rather than necrosis, which occurs when plasma membrane damage is not repaired. Recent reports have described situations in which inflammasome activation and GSDMD cleavage lead to IL-1 β release without cell death²⁷, suggesting that damage to the cell membrane caused by the GSDMD-NT pore can also be repaired under some circumstances. If the double-ring pore exists physiologically, it is possible that under these conditions the second ring could affect whether repair occurs or might even have a signalling function in these activated cells.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0058-6>.

Received: 16 November 2017; Accepted: 8 March 2018;

Published online: 25 April 2018

1. Ding, J. et al. Pore-forming activity and structural autoinhibition of the gasdermin family. *Nature* **535**, 111–116 (2016).
2. Liu, X. et al. Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores. *Nature* **535**, 153–158 (2016).
3. Sborgi, L. et al. GSDMD membrane pore formation constitutes the mechanism of pyroptotic cell death. *EMBO J.* **35**, 1766–1778 (2016).

4. Chen, X. et al. Pyroptosis is driven by non-selective gasdermin-D pore and its morphology is different from MLKL channel-mediated necroptosis. *Cell Res.* **26**, 1007–1020 (2016).
5. Russo, H. M. et al. Active caspase-1 induces plasma membrane pores that precede pyroptotic lysis and are blocked by lanthanides. *J. Immunol.* **197**, 1353–1367 (2016).
6. Shi, J. et al. Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526**, 660–665 (2015).
7. Kayagaki, N. et al. Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* **526**, 666–671 (2015).
8. Aglietti, R. A. et al. GSDMD p30 elicited by caspase-11 during pyroptosis forms pores in membranes. *Proc. Natl Acad. Sci. USA* **113**, 7858–7863 (2016).
9. Rathinam, V. A., Vanaja, S. K. & Fitzgerald, K. A. Regulation of inflammasome signaling. *Nat. Immunol.* **13**, 333–342 (2012).
10. Lamkanfi, M. & Dixit, V. M. Inflammasomes and their roles in health and disease. *Annu. Rev. Cell Dev. Biol.* **28**, 137–161 (2012).
11. Wang, Y. et al. Chemotherapy drugs induce pyroptosis through caspase-3 cleavage of a gasdermin. *Nature* **547**, 99–103 (2017).
12. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
13. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
14. Jacoboni, I., Martelli, P. L., Fariselli, P., De Pinto, V. & Casadio, R. Prediction of the transmembrane regions of β -barrel membrane proteins with a neural network-based predictor. *Protein Sci.* **10**, 779–787 (2001).
15. Murzin, A. G., Lesk, A. M. & Chothia, C. Principles determining the structure of β -sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369–1381 (1994).
16. Law, R. H. et al. The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* **468**, 447–451 (2010).
17. Rossjohn, J., Feil, S. C., McKinstry, W. J., Tweten, R. K. & Parker, M. W. Structure of a cholesterol-binding, thiol-activated cytolytic and a model of its membrane form. *Cell* **89**, 685–692 (1997).
18. Köster, S. et al. Crystal structure of listeriolysin O reveals molecular details of oligomerization and pore formation. *Nat. Commun.* **5**, 3690 (2014).
19. van Pee, K. et al. CryoEM structures of membrane pore and prepore complex reveal cytolytic mechanism of pneumolysin. *eLife* **6**, e23644 (2017).
20. Marshall, J. E. et al. The crystal structure of pneumolysin at 2.0 Å resolution reveals the molecular packing of the pre-pore complex. *Sci. Rep.* **5**, 13293 (2015).
21. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
22. Rosado, C. J. et al. A common fold mediates vertebrate defense and bacterial attack. *Science* **317**, 1548–1551 (2007).
23. Degiacomi, M. T. et al. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat. Chem. Biol.* **9**, 623–629 (2013).
24. Iacovache, I. et al. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nat. Commun.* **7**, 12062 (2016).
25. Idone, V. et al. Repair of injured plasma membrane by rapid Ca²⁺-dependent endocytosis. *J. Cell Biol.* **180**, 905–914 (2008).
26. Keefe, D. et al. Perforin triggers a plasma membrane-repair response that facilitates CTL induction of apoptosis. *Immunity* **23**, 249–262 (2005).
27. Evavold, C. L. et al. The pore-forming protein gasdermin D regulates interleukin-1 secretion from living macrophages. *Immunity* **48**, 35–44 (2018).

Acknowledgements This work was supported by US NIH grant DP1HD087988 (H.W.), R01AI124491 (H.W.), R01AI123265 (J.L.) and Charles A. King Trust Postdoctoral Fellowship Program (J.R., X.L.). We thank the University of Massachusetts Cryo-EM Core Facility and NCI National Cryo-EM Facility at the Frederick National Laboratory for Cancer Research for data collection, and D. Ni and M. Liao for discussions.

Reviewer information Nature thanks H. Saibil, J. Whisstock and B. Zuber for their contribution to the peer review of this work.

Author contributions J.R., H.W. and J.L. conceived the study. J.R., S.X. and H.W. designed the experiments and analysed the data. J.R. reconstituted the pores, performed cryo-EM experiments and refined the structure. J.R. and S.X. analysed structure-based mutants. J.R., H.W. and S.X. analysed the structure. H.W., J.R., S.X. and J.L. wrote the manuscript. X.L. provided advice and comments on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0058-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0058-6>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to H.W. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Cloning, expression and purification of recombinant mouse GSDMA3 and human GSDMD. Full-length mouse GSDMA3 sequence was cloned into a modified pET28a vector with an N-terminal 6× His-SUMO tag, followed by mutagenesis to insert a human rhinovirus 3C protease recognition sequence (LEVLFQGP) immediately after residue E262. The human GSDMD sequence was cloned into the pDB.His.MBP vector to append an N-terminal 6× His-MBP tag followed by a tobacco etch virus (TEV) cleavage site. All mutagenesis was performed using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). All plasmids were verified by DNA sequencing.

To obtain recombinant GSDMA3, *Escherichia coli* BL21 (DE3) cells were transfected with the vector and grown in LB medium supplemented with 50 µg/ml kanamycin. Protein expression was induced overnight at 18 °C with 0.5 mM isopropyl-β-D-thiogalactopyranoside after optical density at 600 nm reached 0.8. Cells were harvested and resuspended in a lysis buffer containing 25 mM Tris-HCl at pH 8.0 and 150 mM NaCl. The protein was first purified by affinity chromatography using Ni-NTA beads (Qiagen) and the 6× His-SUMO tag was removed by overnight Ulp1 protease digestion at 4 °C. The cleaved GSDMA3 was further purified by HiTrap Q ion-exchange chromatography and Superdex 200 gel filtration chromatography (GE Healthcare Life Sciences). Human GSDMD was expressed and purified with similar procedures, except that the N-terminal 6× His-MBP tag was removed by overnight TEV cleavage.

Reconstitution of GSDMA3 and GSDMD pores. 1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and 1', 3'-bis[1,2-dioleoyl-*sn*-glycero-3-phospho]-*sn*-glycerol (sodium salt) (cardiolipin) (Avanti Polar Lipids) dissolved in chloroform were mixed in a glass tube at a mass ratio of 3:1, and the solvent was evaporated under a stream of N₂ gas. Buffer composed of 25 mM Tris-HCl at pH 8.0 and 150 mM NaCl was added to yield a final lipid concentration of 10 mg/ml. Lipid suspension was then vortexed continuously for 5 min. To obtain unilamellar vesicles, liposomes were extruded through a mini-extruder device (Avanti) for 21 passes.

To form GSDMA3 pores on liposomes, purified GSDMA3 was incubated with 3C protease and POPC-cardiolipin liposomes for 4 h on ice. Liposomes were harvested by ultracentrifugation at 60,000g at 4 °C for 30 min and resuspended in 0.5 ml lysis buffer containing 50 mM sodium cholate. After centrifugation at the maximum speed of an Eppendorf centrifuge for 30 min, the supernatant containing the solubilized pores was further purified over a Superose 6 gel-filtration column (GE Healthcare Life Sciences) equilibrated with lysis buffer containing 15 mM sodium cholate. GSDMD pores were formed in the same way but with recombinant caspase-11 in liposomes containing 20% (w/w) phosphatidylserine, and 2% C12E8 was used to solubilize the pores.

Negative staining electron microscopy. For negative staining, 10 µl GSDMA3 or GSDMD pores was placed onto a glow-discharged copper grid (Electron Microscopy Sciences) coated with a layer of thin carbon, washed twice with H₂O, stained with 2% uranyl formate for 40 s and air-dried. The grids were imaged on the Tecnai G² Spirit BioTWIN electron microscope and recorded with an AMT 2k CCD camera (Harvard Medical School Electron Microscopy Facility).

Cryo-EM data collection. A 3-µl drop of native GSDMA3 pores at 2 mg/ml was applied to a glow-discharged Quantifoil grid (R 1.2/1.3 400 mesh, copper, Electron Microscopy Sciences), blotted for 3 s in 100% humidity at 4 °C and plunged into liquid ethane using an FEI Vitrobot Mark IV. The cryo-grids were imaged in an FEI Talos Arctica microscope operating at an acceleration voltage of 200 kV and equipped with a cryo-autoloader (University of Massachusetts Cryo-EM Core Facility). Cryo-EM data were collected automatically on a K2 Summit direct detector camera (Gatan) in super-resolution counting mode, with 8.0 s total exposure time and 200 ms per frame. This resulted in movies each containing 40 frames and an accumulated dose of 41.1 electrons per Å². The super-resolution pixel size is 0.5843 Å. The defocus level in the data collection was set in the range of −1.0 to −3.0 µm. A total of 3,010 movies was collected.

For mercury treatment to label free Cys residues, GSDMA3 pores at 2 mg/ml (~75 µM) were incubated with MgCl₂ at a molar ratio of 1:10 at 4 °C for 2 h. The cryo-grids were plunged in the same way as for the native pores. A total of 1,583 movies was recorded using an FEI Titan Krios electron microscope (National Cryo-electron Microscopy Facility, National Cancer Institute) operating at 300 kV on a K2 Summit direct electron camera operating in super-resolution counting mode, with 12 s total exposure and 300 ms per frame. The defocus range was from −0.75 to −2.0 µm. The resulting movies each contain 40 frames and an accumulated dose of 40.0 electrons per Å². The super-resolution pixel size is 0.66 Å.

Image processing. For data of native GSDMA3 pores, raw movies were corrected by gain reference and for beam-induced motion, and summed into motion-corrected images using MotionCor2²⁸. The CTFFIND4 program was used to

determine the actual defocus level of each micrograph²⁹. We initially selected 462,878 particles from the micrographs with a combination of manual and automatic particle picking in RELION¹². The picked particles were binned two times (pixel size 1.1686 Å) and then subjected to reference-free 2D classification. Good class averages in different orientations were selected for the reconstruction of an initial model in RELION. The resulting initial model, low-pass filtered to 60.0 Å, was used as the input reference to conduct unsupervised 3D classifications in RELION without assumption of any symmetry. The particle set used contained all the side views and top views with apparent C27 symmetry, but not top views with apparent C26 or C28 symmetry. A best 3D class showing symmetric features was selected for further refinement. Upon imposing C27 symmetry, the reconstructions reached a resolution of 4.6 Å, as measured by FSC of a gold-standard refinement in which two halves of the dataset were refined separately and combined only when building the final map¹².

For mercury-soaked GSDMA3 pores, the same procedures were used to process the dataset but two rounds of 3D classification were performed after 2D classification. In brief, 446,553 particles were selected by automatic particle picking in RELION¹². Next, 107,564 particles containing good 2D classes of all side views, top views of C27 and C28 classes, and top views of classes with uncertain symmetry were selected for the first round of 3D classification using an angular sampling angle of 7.5° without applying symmetry. This 3D classification yielded four classes, of which two had apparent C27 and C28 symmetry, respectively. Subsequent 3D classification was performed using an angular sampling angle of 3.7° for both the C27 and C28 classes. From a 3D class of C27-fold symmetry, 40,086 particles were selected for auto-refinement by imposing symmetry in RELION¹². For the C28-fold symmetry pore, 16,581 particles from the 3D class with the best resolution were extracted and refined using SPHIRE³⁰. Final resolutions for C27 and C28 pores are 3.8 Å and 4.2 Å, respectively. Local resolution distribution of all maps was determined by ResMap¹³.

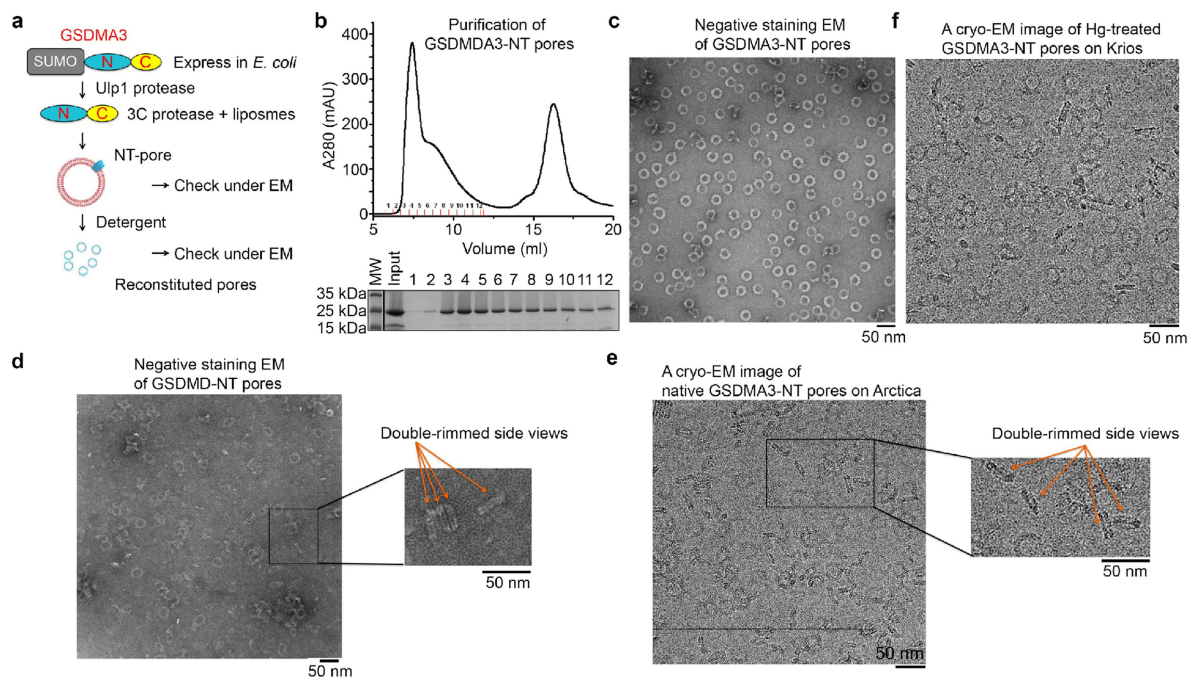
Model building and analysis. The 3.8 Å map with 27-fold symmetry from the mercury-soaked data set and the crystal structure of full length GSDMA3 (PDB ID: 5B5R) were used for model building. The model of the N-terminal domain was docked into the map as a rigid body in Chimera³¹. The fitted model accounted well for most of the density in the globular domain but required extensive remodelling in the β-barrel region using Coot³². PHENIX was used to refine the model against the cryo-EM density in real space and to ensure proper geometry³³. All structural and density representations were generated using either Chimera³¹ or Pymol (<https://www.pymol.org>). Model versus map FSC curves were calculated using Phenix.mtriage to obtain an estimate of the final resolution for the models.

Liposome leakage assay. The leakage of liposomes encapsulating TbCl₃ was determined by an increase in fluorescence intensity when Tb³⁺ bound to dipicolinic acid (DPA) in the external buffer. Tb³⁺-entrapped liposomes were prepared using a similar procedure as detailed in 'Reconstitution of GSDMA3 and GSDMD pores', with the exception that the buffer contained 20 mM HEPES at pH 7.4, 150 mM NaCl, 50 mM sodium citrate and 15 mM TbCl₃. The suspension was loaded onto a Superose 6 gel filtration column to remove unincorporated TbCl₃. Purified liposomes were supplemented with 50 µM DPA before addition of a recombinant GSDM protein (0.5 µM) and caspase-11 or 3C protease (0.2 µM). Fluorescence at 545 nm after excitation at 276 nm was recorded continuously for 45 min at 30-s intervals using a Molecular Devices SpectraMax M5 plate reader.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-7449 (HgCl₂-treated, C27 membrane pore), EMD-7450 (HgCl₂-treated, C28 membrane pore) and EMD7451 (native, double-ring pore). The atomic structure coordinates have been deposited in the Protein Data Bank under the accession number 6CB8. All other data can be obtained from the corresponding author upon reasonable request.

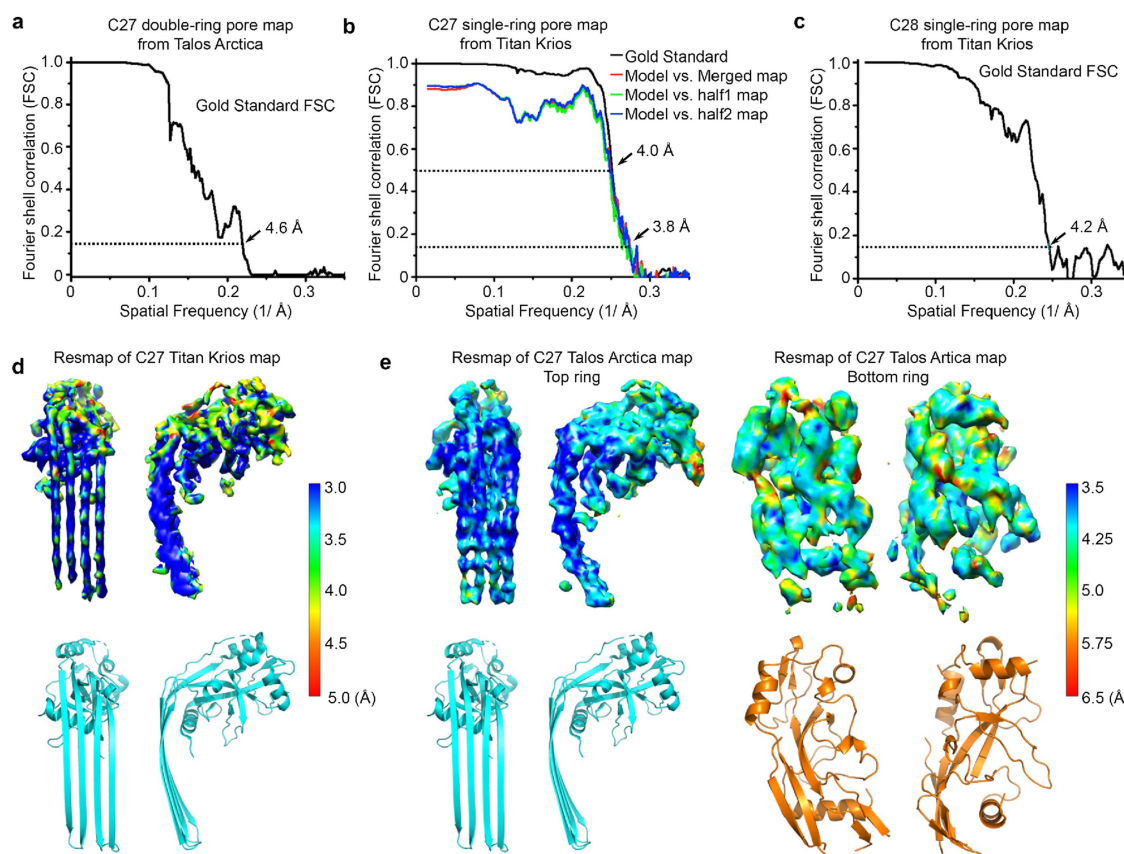
28. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
29. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
30. Moriya, T. et al. High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.* 123, <https://doi.org/10.3791/55448> (2017).
31. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
33. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).



Extended Data Fig. 1 | Reconstitution of GSDMA3 pores in vitro.

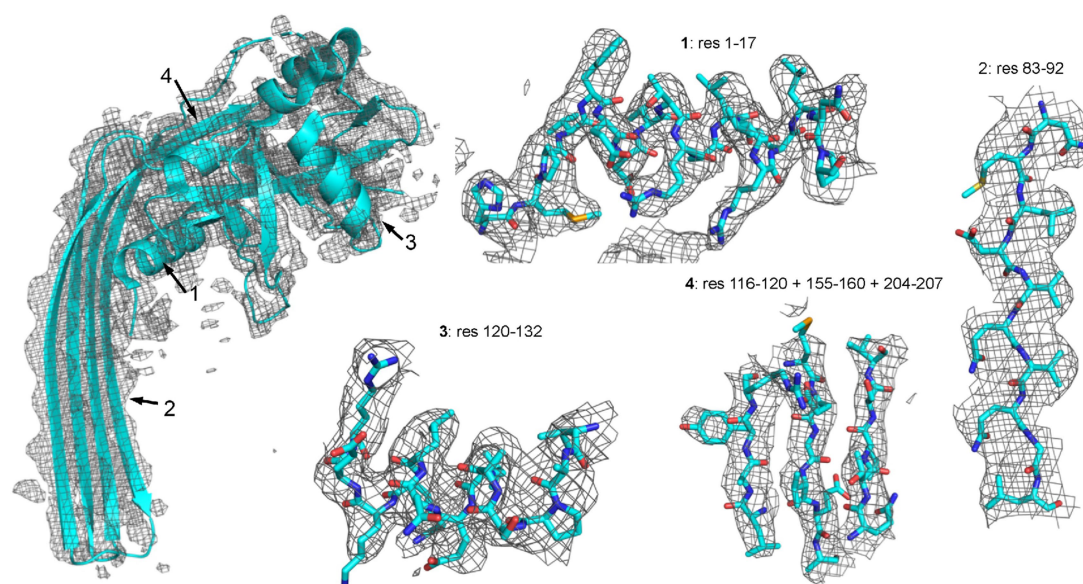
a, A general procedure for GSDMA3 pore reconstitution. **b**, Size-exclusion chromatography of GSDMA3 pores extracted from liposomes (top) and Coomassie blue-stained SDS-PAGE of the collected fractions. **c**, Representative negative-stain electron microscopy images of GSDMA3.

d, Representative negative-stain electron microscopy images of human GSDMD pores with double-rimmed side views. **e**, **f**, Representative cryo-EM images of GSDMA3 pores with double-rimmed side views (**e**) and HgCl₂-treated GSDMA3 pores (**f**). Scale bar, 50 nm (**c**–**f**). All results were confirmed at least three times as technical replicates.

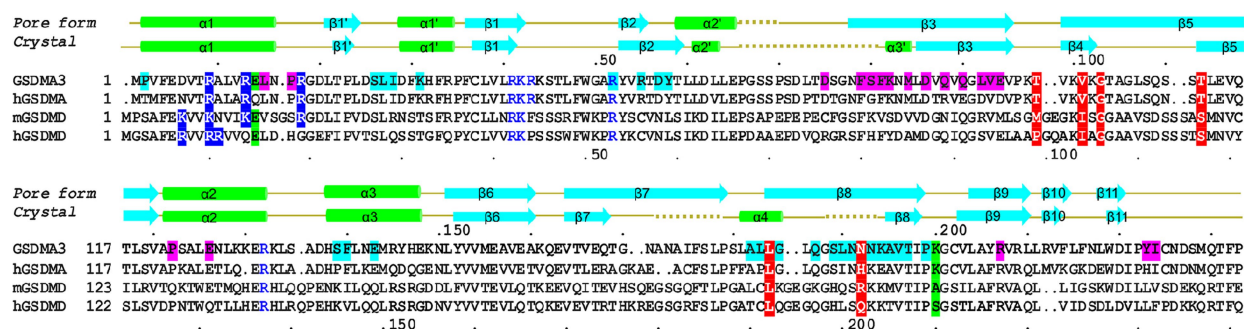


Extended Data Fig. 2 | Cryo-EM analysis of double- and single-ring GSDMA3 pores. **a–c**, Gold-standard FSC plots from two half-reconstructions refined separately in RELION for the 27-fold double-ring pore map (**a**), the 27-fold single-ring pore map (**b**) and the 28-fold single-ring pore map (**c**). Model-to-map correlations are shown for the 27-fold single-ring pore (**b**). **d, e**, Local resolution estimation generated

by ResMap on the two half maps separately refined in a gold-standard procedure in RELION. Local resolutions are colour-coded on the densities. Highest resolution is observed at the β -barrel domain for both the single-ring pore (**d**) and the membrane-inserted ring of the double-ring pore (**e**). The globular domains and the additional ring exhibit relatively low resolution.

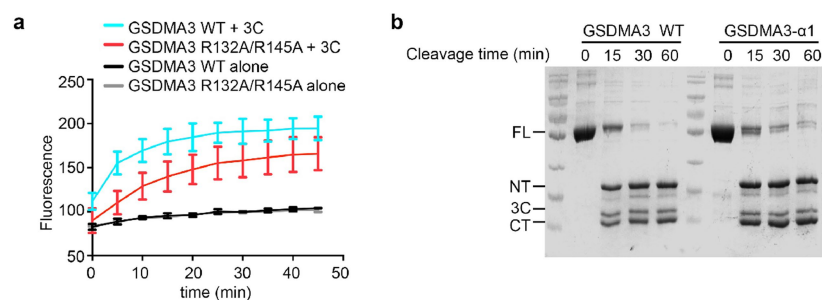


Extended Data Fig. 3 | Cryo-EM density validation. Close-up views of GDSMA3 subunit model fitted into the cryo-EM density map at four locations denoted by residue numbers.



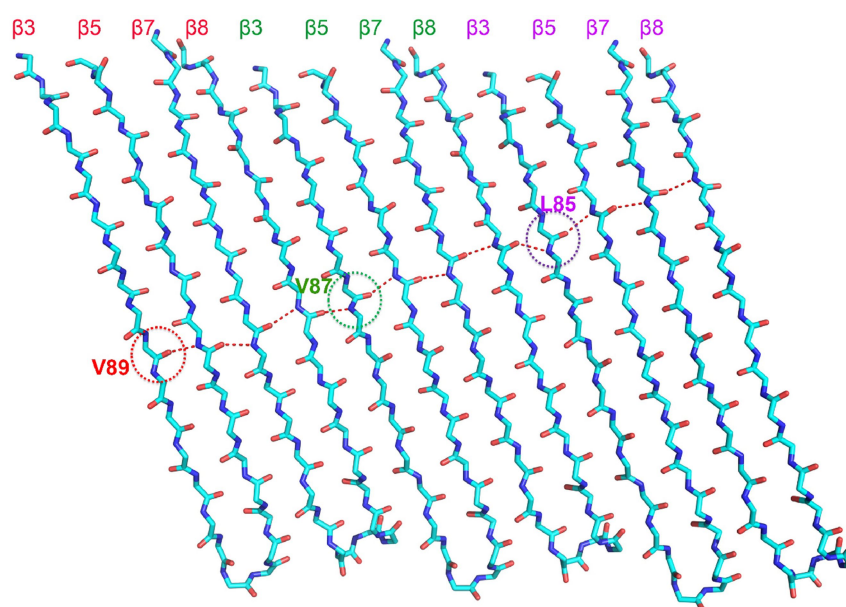
Extended Data Fig. 4 | Aligned sequences in the N-terminal region among GSDMA3, human GSDMA, mouse GSDMD and human GSDMD. Secondary structures are shown for both the auto-inhibited conformation¹ and the pore conformation of GSDMA3. Dotted lines represent disordered regions. Residue numbers are shown as dots every ten residues above the alignment for GSDMA3 and below for human GSDMD. Blue highlights, potential positively charged residues for acidic

lipid binding; blue, tested residues that did not affect pore formation; magenta and cyan highlights, residues involved in the oligomerization or membrane insertion of the two adjacent subunits; red and green highlights, previously reported and cancer-associated mutations localized on the oligomerization interface of the globular domains (green) and β -barrel domain (red).

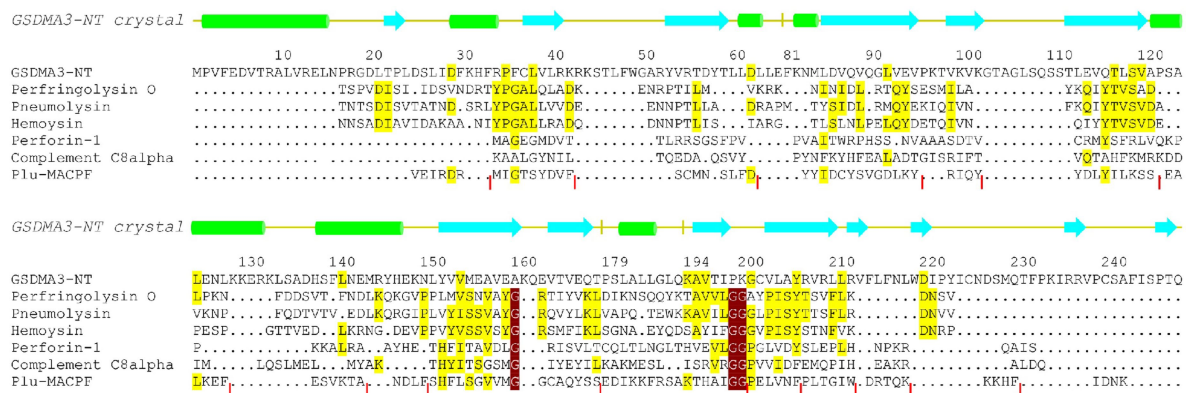


Extended Data Fig. 5 | Lipid binding by GSDM. **a**, Effect of the GSDMA3 R132A/R145A mutation on the liposome-leakage activity, monitored by measuring DPA-chelating-induced fluorescence of released Tb^{3+}

ion ($n = 3$ biological replicates). Error bars denote mean \pm s.d. **b**, Equal processing of GSDMA3 by the 3C protease for the wild type and the α 1 mutant. A representative gel of two independent experiments is shown.



Extended Data Fig. 6 | Two-residue shift in the hydrogen-bonding pattern in each GSDMA3 subunit. The shift results in a shear number of $27 \times 2 = 54$ for a 27-fold symmetric pore.



Extended Data Fig. 7 | Structure-based sequence alignment. The N-terminal region of GSDMA3 crystal structure (PDB ID: 5B5R) aligned with structures of perfringolysin O (5DIM), pneumolysin (4QQQ), haemolysin (3HVN), perforin-1 (3NSJ), complement C8 alpha (2QQH) and Plu-MAPCF (2QP2) and using the distance alignment matrix method (DALI). Helices are coloured in green, strands in cyan and loops in yellow.

Vertical yellow bars indicate disordered regions in the GSDMA3-NT structure. Vertical red bars denote gaps in alignment of Plu-MAPCF. Identical residues conserved among at least three proteins are highlighted in yellow. Glycines conserved among MACPF/CDCs are not conserved in GSDMA3, as highlighted in deep red.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	#1 Hg-treated, C27-symmetry (EMDB-7449) (PDB 6CB8)	#2 Hg-treated, C28-symmetry (EMDB-7450)	#3 Native, C27-symmetry Stacked double pore (EMDB-7451)
Data collection and processing			
Voltage (kV)	300	300	200
Electron exposure (e-/Å ²)	40.0	40.0	41.1
Defocus range (μm)	-0.75 to -2.0	-0.75 to -2.0	-0.1 to -3.0
Super-resolution pixel size (Å)	0.66	0.66	0.5843
Symmetry imposed	C27	C28	C27
Initial particle images (no.)	446,553	446,553	462,878
Final particle images (no.)	40,086	16,581	64,248
Map resolution (Å)	3.8	4.2	4.6
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	60.0-3.8	60.0-4.2	60.0-4.6
Refinement			
Initial model used (PDB code)	5B5R		
Model resolution (Å)	4.0		
FSC threshold	0.5		
Model resolution range (Å)	84.5-3.8		
Map sharpening <i>B</i> factor (Å ²)	-131.255		
Model composition			
Non-hydrogen atoms	1,844		
Protein residues	230		
Ligands	14		
<i>B</i> factors (Å ²)			
Protein	101.86		
Ligand	92.69		
R.m.s. deviations			
Bond lengths (Å)	0.009		
Bond angles (°)	1.122		
Validation			
MolProbity score	2.46 (99th percentile)*		
Clashscore	16.22 (97th percentile)*		
Poor rotamers (%)	0.00		
Ramachandran plot			
Favored (%)	86.28		
Allowed (%)	13.72		
Disallowed (%)	0.00		

*100th is the best among the structures of comparable resolution, 0th is the worst.

Helium in the eroding atmosphere of an exoplanet

J. J. Spake^{1*}, D. K. Sing^{1,2}, T. M. Evans¹, A. Oklopčić³, V. Bourrier⁴, L. Kreidberg^{5,6}, B. V. Rackham⁷, J. Irwin⁶, D. Ehrenreich⁴, A. Wyttenbach⁴, H. R. Wakeford⁸, Y. Zhou⁷, K. L. Chubb⁹, N. Nikolov¹, J. M. Goyal¹, G. W. Henry¹⁰, M. H. Williamson¹⁰, S. Blumenthal¹, D. R. Anderson¹¹, C. Hellier¹¹, D. Charbonneau⁶, S. Udry⁴ & N. Madhusudhan¹²

Helium is the second-most abundant element in the Universe after hydrogen and is one of the main constituents of gas-giant planets in our Solar System. Early theoretical models predicted helium to be among the most readily detectable species in the atmospheres of exoplanets, especially in extended and escaping atmospheres¹. Searches for helium, however, have hitherto been unsuccessful². Here we report observations of helium on an exoplanet, at a confidence level of 4.5 standard deviations. We measured the near-infrared transmission spectrum of the warm gas giant³ WASP-107b and identified the narrow absorption feature of excited metastable helium at 10,833 angstroms. The amplitude of the feature, in transit depth, is 0.049 ± 0.011 per cent in a bandpass of 98 angstroms, which is more than five times greater than what could be caused by nominal stellar chromospheric activity. This large absorption signal suggests that WASP-107b has an extended atmosphere that is eroding at a total rate of 10^{10} to 3×10^{11} grams per second (0.1–4 per cent of its total mass per billion years), and may have a comet-like tail of gas shaped by radiation pressure.

WASP-107b is one of the lowest-density planets known, with a radius ($0.94 \pm 0.02 R_J$) similar to that of Jupiter but a much lower mass³, ($0.12 \pm 0.01 M_J$); R_J and M_J are the radius and mass of Jupiter, respectively. It orbits an active K6 dwarf every 5.7 days at a distance of 0.055 ± 0.001 astronomical units. On 31 May 2017, we observed a primary transit of WASP-107b with Wide Field Camera 3 (WFC3), which is onboard the Hubble Space Telescope (HST). Our observations lasted 7 h and we acquired 84 time-series spectra with the G102 grism, which covers the 8,000–11,000 Å wavelength range. Further details about the observations and data reduction can be found in Methods.

Each spectrum was integrated along the wavelength axis to first produce a ‘white’ light curve (Extended Data Fig. 1). In addition to the planetary transit signal, the resulting time series was affected by instrumental systematic errors caused by electron trapping in the WFC3 detector⁵. We fitted the white-light curve with a planetary transit model⁴ multiplied by a linear baseline trend and a physically motivated WFC3 systematics model⁵. For the planetary transit model, we allowed the planet-to-star radius ratio (R_p/R_s) and the mid-transit time (t_0) to vary as a free parameter, while holding the ratio of the orbital distance to the stellar radius (a/R_s), the inclination (i), the eccentricity (e) and the period (P) fixed to previously determined values^{6,7}. We assumed a quadratic limb-darkening profile for the star, holding the coefficients fixed to values determined from a model stellar spectrum⁸. Further details about this fit are provided in Methods. The results of the fit are reported in Extended Data Table 1 and Extended Data Fig. 1.

Two sets of spectroscopic light curves were constructed by summing each spectrum into broad- and narrowband bins. The first set consisted of 9 broadband channels spanning the 8,770–11,360 Å wavelength range, and the second set comprised 20 overlapping, narrowband channels spanning the 10,580–11,070 Å wavelength range. The narrowband channels covered the helium absorption triplet at 10,833 Å (vacuum

wavelength; the air wavelength of this line is 10,830 Å). The widths of the broadband and narrowband channels were 294 Å (12-pixel columns) and 98 Å (4-pixel columns), respectively. We fitted both sets of spectroscopic light curves using the approach described above for the white-light curve. However, for the planetary transit signals, we only allowed R_p/R_s to vary as a free parameter, while holding t_0 , a/R_s , i , e and P fixed to the values reported in Extended Data Table 1. We fixed the limb-darkening coefficients in a similar way to the white-light curve fit. Additional details of the fitting procedure are given in Methods. The inferred values for the transit depth, (R_p/R_s)², in each wavelength channel are shown in Fig. 1 and Extended Data Table 2. These results constitute the atmospheric transmission spectrum.

The broadband transmission spectrum is consistent with a previous transmission spectrum for WASP-107b, obtained using the WFC3 G141 grism, which covers the 11,000–16,000 Å wavelength range⁹. The latter exhibits a muted water absorption band centred at 14,000 Å, with an otherwise flat spectrum implying an opaque cloud deck. After applying a correction for stellar activity variations between the G102 and G141 observation epochs (see Methods), the G102 spectrum aligns with the cloud deck level inferred from the G141 spectrum (Fig. 1).

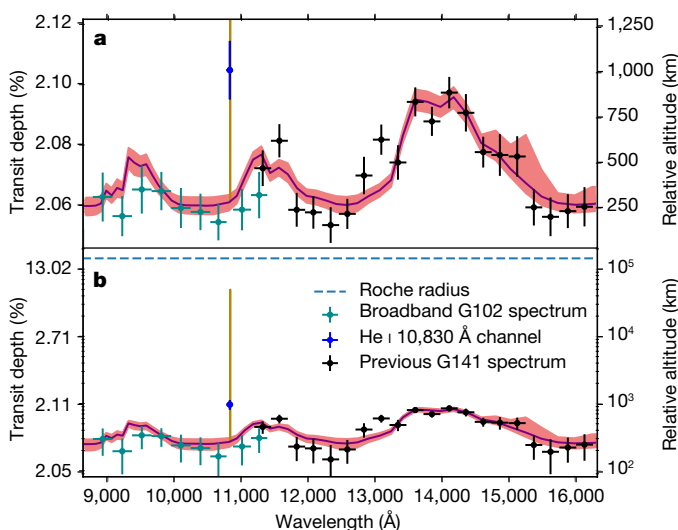


Fig. 1 | Combined near-infrared transmission spectrum for WASP-107b with the helium absorption feature. **a**, Data are plotted on a linear scale. Points, with 1σ error bars, are from a previous study (black)⁹ and this work (light and dark blue), both corrected for stellar activity (see Methods). The solid purple line is the best fit obtained with a lower-atmosphere retrieval model based on the Markov chain Monte Carlo technique, and the pink shaded area encompasses 99.7% of the Markov chain Monte Carlo samples. The gold line is the best-fitting absorption profile for the 10,830-Å helium line, obtained by our 1D escaping-atmosphere model. **b**, Same as **a**, on a logarithmic scale. The dashed blue line shows the Roche radius.

¹Department of Physics and Astronomy, University of Exeter, Exeter, UK. ²Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA. ³Institute for Theory and Computation, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA. ⁴Observatoire de l'Université de Genève, 51 chemin des Maillettes, Versoix, Switzerland. ⁵Harvard Society of Fellows, Cambridge, MA, USA. ⁶Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA. ⁷Department of Astronomy/Steward Observatory, University of Arizona, Tucson, AZ, USA. ⁸Space Telescope Science Institute, Baltimore, MD, USA. ⁹Department of Physics and Astronomy, University College London, London, UK. ¹⁰Center of Excellence in Information Systems, Tennessee State University, Nashville, TN, USA. ¹¹Astrophysics Group, Keele University, Keele, UK. ¹²Institute of Astronomy, University of Cambridge, Cambridge, UK. *e-mail: jspake@astro.ex.ac.uk

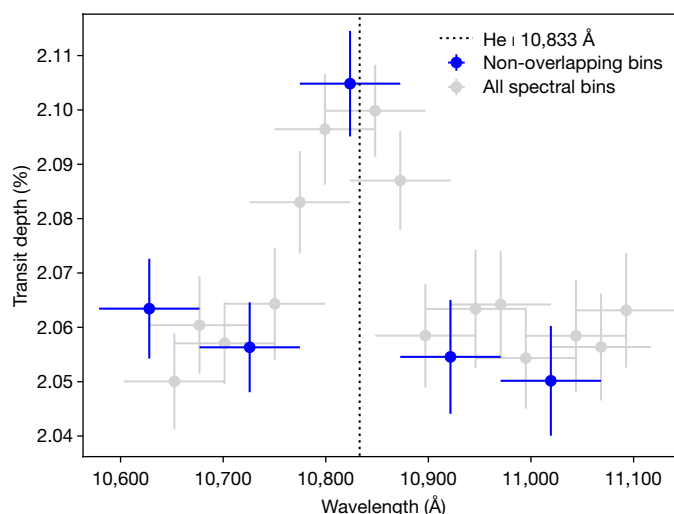


Fig. 2 | Narrowband transmission spectrum of WASP-107b, centred on 10,833 Å. Each spectroscopic channel has been shifted one pixel away from the next one. Non-overlapping bins are shown by blue symbols. All error bars are 1σ . The peak of the spectrum coincides with the 2^3S helium absorption line at 10,833 Å.

The helium triplet has an expected width of approximately 3 Å, whereas the resolution of the G102 grism is 67 Å (about 3 pixels)¹⁰ at 10,400 Å. Therefore, to make a finely sampled transmission spectrum, we shifted each of the 20 narrowband channels by one pixel with respect to the adjacent channel along the wavelength axis. The narrowband transmission spectrum peaked at the channel most closely centred at 10,833 Å (Fig. 2), as would be expected if helium absorption in the planetary atmosphere were responsible for the signal. To estimate the amplitude of the absorption feature, we focused on five non-overlapping channels centred on 10,833 Å. All but one of these channels were consistent with a baseline transit-depth level of $2.056\% \pm 0.005\%$. The single exception was the channel centred on the 10,833-Å helium triplet, where the transit depth was visibly larger than for the surrounding channels (Fig. 3), and we obtained $(R_p/R_s)^2 = 2.105\% \pm 0.010\%$. We ruled out various alternative explanations for the signal, including other absorbing species, helium in Earth's atmosphere, and occultation of inhomogeneities in the stellar chromosphere and photosphere (see Methods).

The metastable helium probed by 10,833-Å line absorption forms high up in planetary atmospheres, at microbar- to nanobar-level pressures, where stellar extreme-ultraviolet radiation is absorbed¹¹. On the other hand, absorption of the neighbouring continuum occurs deeper in planetary atmospheres, at millibar- to bar-level pressures. Therefore, to interpret the broadband (continuum) and narrowband (about 10,833 Å) transmission spectra, we used separate lower- and upper-atmosphere models. For the combined G102 and G141 broadband spectrum (with the 10,775–10,873 Å range removed), we performed an atmospheric retrieval analysis using our one-dimensional (1D) radiative transfer code, ATMO^{12,13} (see Methods and Extended Data Table 3). We found that the broadband data were well explained by a grey absorbing cloud deck across the full 8,780–11,370 Å wavelength range, in addition to H₂O absorption. We obtained a volume mixing ratio of 5×10^{-3} – 4×10^{-2} for H₂O, consistent with previous estimations⁹.

We investigated the narrowband transmission spectrum using two numerical models for the upper atmosphere of WASP-107b (see Methods). The first, a 1D model¹⁴, solves for the level populations of a H/He Parker wind, and suggests that WASP-107b is losing its atmosphere at a rate of 10^{10} – 3×10^{11} g s⁻¹, corresponding to about 0.1%–4% of its total mass every billion years. The second, a three-dimensional (3D) model^{16,77} obtains an escape rate of 10^6 – 10^7 g s⁻¹ for metastable helium (for comparison, the 1D model gives an escape rate of about 10^5 g s⁻¹ for 2^3S helium). It also suggests that stellar radiation

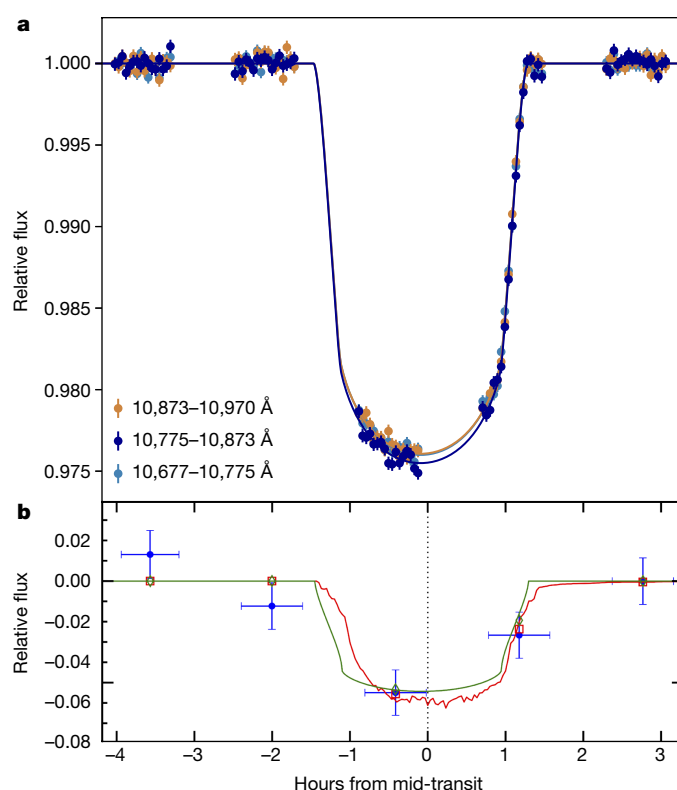


Fig. 3 | Transit light curves for three 98-Å-wide spectroscopic channels.

a, Dark-blue points are from the channel centred on the He I 10,833-Å line; gold and light-blue points are from the two adjacent channels. All data points have 1σ error bars. Solid lines are best-fit light-curve model results. The transit depth of the dark-blue curve is visibly the largest. **b**, Binned difference between the light curves of the 10,775-Å and 10,873-Å channels, and the average of the two adjacent channels (blue points, 1σ errors), highlighting the excess absorption. This difference is well explained by both our 1D (green line) and 3D (red line) escaping-atmosphere models. Green and red points are binned model results.

pressure blows the escaping helium atoms away from the planet so swiftly that they form a tail nearly aligned with the star–planet axis; this could explain the lack of post-transit occultation detected in our data (Fig. 3). The radiation pressure may also blue-shift the absorption signature over velocities of hundreds of kilometres per second, which would be observable at higher spectral resolution (Fig. 4).

Atmospheric mass loss can substantially alter the bulk composition of a planet. For example, there is evidence that atmospheric escape is responsible for the observed dearth of highly irradiated super-Earth and sub-Neptune exoplanets with sizes between 1.6 and 2 Earth radii^{17–21}. To evaluate planet formation theories and assess whether these planets have substantial H/He envelopes, it is necessary to understand how atmospheric mass loss affects the subsequent evolution of bodies that start with sizeable atmospheres. Empirical constraints, such as that presented here for WASP-107b, are therefore crucial for retracing evolutionary pathways and interpreting the present-day population of planets²².

Until now, extended atmospheres have been detected on three exoplanets by targeting the Ly α line in the ultraviolet^{15,23,24} and on one exoplanet using the optical H α line²⁵. Our observations of WASP-107b using the 10,833-Å line provide not only the first detection of helium on an exoplanet, but also the first detection of an extended exoplanet atmosphere at infrared wavelengths. Our results demonstrate the feasibility of a new method for studying extended atmospheres that is complementary to those using the two hydrogen lines.

We note that ground observations targeting the 10,833-Å helium triplet are possible with existing high-resolution infrared spectrographs. In the near future, high-signal-to-noise observations will also

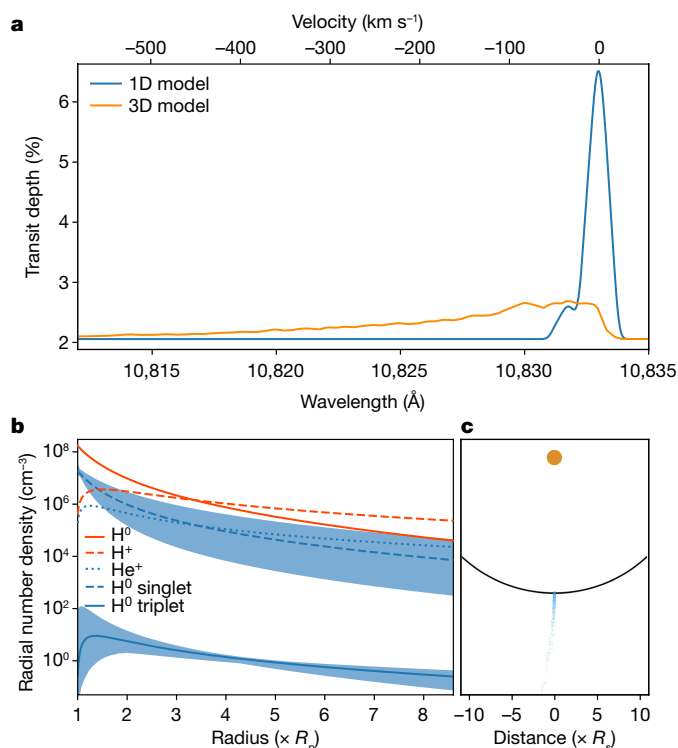


Fig. 4 | Results of the two models for the upper atmosphere of WASP-107b. **a**, Best-fitting absorption profiles of the helium 10,833-Å triplet line from the 1D (blue), and 3D (orange) models. Both models reproduce the measured excess absorption of $0.049\% \pm 0.011\%$ in a 98-Å bin. Higher-resolution observations will resolve the profile shape and further constrain the velocity of the planetary wind. **b**, Radial number density profiles of different atmospheric species from the 1D model; blue-shaded regions are 1σ errors. **c**, Top-down view of the planetary system from the 3D model, showing a comet-like tail of 2³S helium shaped by stellar radiation pressure.

be possible with the James Webb Space Telescope at a spectral resolution of about 4 Å (approximately 110 km⁻¹).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0067-5>.

Received: 5 December 2017; Accepted: 9 March 2018;

Published online 2 May 2018.

- Seager, S. & Sasselov, D. D. Theoretical transmission spectra during extrasolar giant planet transits. *Astrophys. J.* **537**, 916–921 (2000).
- Moutou, C., Coustenis, A., Schneider, J., Queloz, D. & Mayor, M. Searching for helium in the exosphere of HD 209458b. *Astron. Astrophys.* **405**, 341–348 (2003).
- Anderson, D. et al. The discoveries of WASP-91b, WASP-105b and WASP-107b: two warm Jupiters and a planet in the transition region between ice giants and gas giants. *Astron. Astrophys.* **604**, A110 (2017).
- Kreidberg, L. batman: Basic Transit Model cAlculationN in Python. *Publ. Astron. Soc. Pacif.* **127**, 1161 (2015).
- Zhou, Y., Apai, D., Lew, B. W. P. & Schneider, G. A physical model-based correction for charge traps in the Hubble Space Telescope's Wide Field Camera 3 near-IR detector and its applications to transiting exoplanets and brown dwarfs. *Astron. J.* **153**, 243 (2017).
- Fossati, L. et al. Metals in the exosphere of the highly irradiated planet WASP-12b. *Astrophys. J.* **714**, L222–L227 (2010).
- Dai, F. & Winn, J. N. The oblique orbit of WASP-107b from K2 photometry. *Astron. J.* **153**, 205 (2017).
- Castelli, F. & Kurucz, R. L. New grids of ATLAS9 model atmospheres. Preprint at <https://arxiv.org/abs/astro-ph/0405087> (2004).
- Kreidberg, L., Line, M., Thorngren, D., Morley, C. & Stevenson, S. Water, methane depletion, and high-altitude condensates in the atmosphere of the warm super-Neptune WASP-107b. Preprint at <https://arxiv.org/abs/1709.08635> (2018).

- Kuntzschner, H., Bushouse, H., Kummel, M. & Walsh, J. R. *WFC3 SMOV Proposal 11552: Calibration of the G102 Grism*. Report No. WFC3–2009–18 (Space Telescope European Coordinating Facility, 2009); <http://www.stsci.edu/hst/wfc3/documents/ISRs/WFC3-2009-18.pdf>.
- Christie, D., Arras, P. & Li, Z. H α absorption in transiting exoplanet atmospheres. *Astrophys. J.* **772**, 144 (2013).
- Amundsen, D. et al. Accuracy tests of radiation schemes used in hot Jupiter global circulation models. *Astron. Astrophys.* **564**, A59 (2014).
- Tremblin, P. et al. Fingering convection and cloudless models for cool brown dwarf atmospheres. *Astrophys. J.* **804**, L17 (2015).
- Oklopčić, A. & Hirata, C. M. A new window into escaping exoplanet atmospheres: 10830 Å line of metastable helium. *Astrophys. J.* **855**, L11 (2018).
- Kulow, J. R., France, K., Linsky, J. & Loyd, R. O. P. Ly α transit spectroscopy and the neutral hydrogen tail of the hot Neptune GJ 436b. *Astrophys. J.* **786**, 132 (2014).
- Bourrier, V., Lecavelier des Etangs, A., Ehrenreich, D., Tanaka, Y. A. & Vidotto, A. A. An evaporating planet in the wind: stellar wind interactions with the radiatively braked exosphere of GJ 436 b. *Astron. Astrophys.* **591**, A121 (2016).
- Lopez, E. D., Fortney, J. J. & Miller, N. How thermal evolution and mass-loss sculpt populations of super-Earths and sub-Neptunes: application to the Kepler-11 system and beyond. *Astrophys. J.* **761**, 59 (2012).
- Owen, J. & Wu, Y. Kepler planets: a tale of evaporation. *Astrophys. J.* **775**, 105 (2013).
- Jin, S. et al. Planetary population synthesis coupled with atmospheric escape: a statistical view of evaporation. *Astrophys. J.* **795**, 65 (2014).
- Chen, H. & Rogers, L. A. Evolutionary analysis of gaseous sub-Neptune-mass planets with MESA. *Astrophys. J.* **831**, 180 (2016).
- Fulton, B. et al. The California-Kepler survey. III. A gap in the radius distribution of small planets. *Astron. J.* **154**, 109 (2017).
- Lopez, E. & Fortney, J. J. Understanding the mass-radius relation for sub-Neptunes: radius as a proxy for composition. *Astrophys. J.* **792**, 1 (2014).
- Vidal-Madjar, A. et al. An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* **422**, 143–146 (2003).
- Lecavelier des Etangs, A. et al. Evaporation of the planet HD 189733b observed in H I Lyman- α . *Astron. Astrophys.* **514**, A72 (2010).
- Jensen, A. G. et al. A detection of H α in an exoplanetary exosphere. *Astrophys. J.* **751**, 86 (2012).

Acknowledgements We thank S. Seager, A. Dupree, V. Andretta, M. Giampapa and B. Drummond for discussions. This work is based on observations made with the NASA/ESA HST, obtained at the Space Telescope Science Institute (STScI) operated by AURA, Inc. J.J.S. is supported by an STFC studentship. The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement number 336792. Support for this work was provided by NASA grants under the HST-GO-14916 programme of the STScI. G.W.H. and M.H.W. acknowledge support from Tennessee State University and the State of Tennessee through its Centers of Excellence programme. The MEarth Team (J.J. and D.C.) gratefully acknowledges funding from the David and Lucille Packard Fellowship for Science and Engineering, the US National Science Foundation (NSF) and the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. This work was carried out in the framework of the National Centre for Competence in Research PlanetS supported by the Swiss National Science Foundation (SNSF). V.B., D.E., A.W. and S.U. acknowledge financial support from the SNSF. D.E. and V.B. acknowledge funding from the ERC under the European Union's Horizon 2020 research and innovation programme (project FOUR ACES; grant agreement No 724427). B.V.R. acknowledges support from an NSF Graduate Research Fellowship (grant DGE-1746060) and the Earth and Other Solar Systems Team, NASA Nexus for Exoplanet System Science. IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the NSF. J.M.G. acknowledges support from a Leverhulme Trust Research Project Grant.

Reviewer information Nature thanks D. Deming and S. Redfield for their contribution to the peer review of this work.

Author contributions J.J.S. led the HST time proposal, designed the observations, and led the data analysis with contributions from T.M.E., H.R.W., L.K. and Y.Z. J.J.S. identified the planetary helium and wrote the manuscript with contributions from T.M.E., V.B., A.O., J.I., B.V.R. and G.W.H. A.O. and V.B. performed detailed modelling of the exosphere, with contributions from D.E. D.K.S. provided scientific guidance and performed the retrieval analysis. J.I., G.W.H., M.H.W. and D.C. provided ground-based photometry to correct for stellar activity. All authors discussed the results and commented on the paper.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0067-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.J.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Observations and data reduction. We observed one transit of WASP-107b with WFC3 in spectroscopic mode using the G102 grism (programme GO-14916; principal investigator, J.J.S.), which covers the approximate wavelength range of 8,780–11,370 Å. We used forward spatial scanning to spread the spectra over about 60 pixels in the cross-dispersion direction with the SPARS10, NSAMP = 15 setup, giving exposure times of about 103 s. This allowed 17 exposures per HST orbit. The observations lasted for five HST orbits, with two orbits pre-transit, one during the transit and one post-transit, allowing us to constrain the out-of-transit baseline precisely.

The raw frames were automatically reduced with the CalWF3 pipeline (version 3.3)²⁶ of STScI. From the reduced frames we extracted the 1D spectra following standard methods²⁷: we built up flux counts by summing the difference between successive non-destructive reads. We then removed the background from each read difference by subtracting the median of a box of pixels uncontaminated by the spectrum. We found the flux-weighted centre of each scan and set to zero all pixels that were more than 75 rows away from the centre in the cross-dispersion axis, which removed many cosmic rays. The remaining cosmic rays were flagged by finding 4σ outliers relative to the median along the dispersion direction. We replaced each flagged pixel with the median along the dispersion direction, which was rescaled to the count rate of the cross-dispersion column. Because the scans were visibly tilted from the dispersion axis, we used the package Apall of the Image Reduction and Analysis Facility to fit the trace of the two-dimensional scans and extract 1D spectra. We found the wavelength solutions by cross-correlating the extracted spectra with the ATLAS model stellar spectrum⁸ that most closely matches WASP-107 (effective temperature, $T_{\text{eff}} = 4,500$ K; surface gravity, $\log(g) = 4.5$ in CGS units), modulated by the G102 grism throughput. Following standard methods⁹ we interpolated each spectrum onto the wavelength range of the first exposure to account for shifts in the dispersion axis over time.

White-light curve analysis. We extracted the white-light curve by summing the total counts of each 1D spectrum. To constrain the mid-time of the transit, we fitted the resulting time series with the BATMAN transit model⁴, multiplied by a linear baseline trend and a physically motivated systematics model. For the latter, we employed the RECTE model³, which accounts for two populations of charge traps in individual pixels of the detector and successfully replicates the ramp-like features that dominate the systematics. The RECTE model allows us to keep the first-orbit observations in our fit. The free parameters of our final model were: the planet-to-star radius ratio, R_p/R_s ; the mid-transit time, t_0 ; the gradient (c_1) and y-intercept (c_0) of the linear background trend; four parameters for the charge-trapping model—the initial numbers of populated slow (s_{pop}) and fast (f_{pop}) traps, and the corresponding changes in the two populations (δ_s and δ_f) between each orbit; and an uncertainty rescaling factor, β , for the expected photon noise. We fixed a/R_s , i , e , and the period using estimates from Kepler light curves⁷. To model the stellar limb darkening, we fitted a four-parameter nonlinear limb-darkening law²⁸ to the ATLAS stellar model described above.

Because the shape of the ramp-like systematics depends on the count level of the illuminated pixels, the RECTE model requires the ‘intrinsic’ count rate of a pixel (that is, the actual flux received from the star) to model the charge trapping. To create a template of the intrinsic count rate, we median-combined four raw images from the end of the second orbit, in which the charge traps appear completely filled and the ramp shape of the light curve has tapered to a flat line. Although it is possible to model each illuminated pixel, this is computationally expensive for a large scan. Additionally, the ramp profile is blurred by systematic errors introduced by telescope jittering and pointing drift. Instead, we divided the scan into columns with a width of 10 pixels along the dispersion axis and fed the median-count profiles into the model.

We used the Markov chain Monte Carlo (MCMC) package emcee²⁹ to marginalize over the parameter space of the model likelihood distribution. We used 80 walkers and ran chains for 8,000 steps, discarding the first 800 as burn-in, before combining the walker chains into a single chain. The best-fitting model and residuals are shown in Extended Data Fig. 1, with the parameter values and 1σ uncertainties reported in Extended Data Table 1. Although WASP-107b orbits an active star, we see no evidence of star spot crossings. For context, only five spot-crossing events have been reported in 10 Kepler transits^{7,30}.

Broadband spectroscopic light-curve fit. We binned each spectrum into nine spectroscopic channels across the 8,780–11,370 Å wavelength range, each spanning 10–12 detector pixels. The resulting light curves are shown in Extended Data Fig. 1. Because the throughput of the G102 grism is wavelength-dependent, the shape of the charge-trapping ramp in each spectroscopic light curve is different. Therefore, we simultaneously fitted each channel using a transit model multiplied by a linear baseline trend and a charge-trap model. To make a template of the intrinsic counts, we took the median cross-dispersion-direction profile of each channel in the same four raw images as used in the white-light curve fit. We fixed t_0 to the value found from the white-light curve fit. Similarly to the white-light

curve fit, we fixed the orbital parameters to those derived from Kepler light curves⁷ and wavelength-dependent limb-darkening coefficients from the ATLAS model. Therefore, for each channel the fitted parameters were R_p/R_s , c_1 , c_0 , s_{pop} , f_{pop} , δ_s , δ_f , and β . We ran MCMC fits for each light curve using emcee with 80 walkers, 80,000 steps and a burn-in of 800.

As a test, we ran additional fits for the spectroscopic light curves with the stellar limb-darkening coefficients as free parameters. This produced results that were consistent, within 1σ , with those obtained from the analysis in which the limb-darkening coefficients were fixed.

We show the resulting spectroscopic light curves divided by their best-fitting systematics models in Extended Data Fig. 1, along with their residuals. Extended Data Table 2 reports our median values for the transit depth, $(R_p/R_s)^2$, with 1σ uncertainties calculated from the MCMC chains. We also list the root mean square of the residuals for each channel, which range between 1.038–1.198 times the photon noise.

Narrowband spectroscopic light-curve fit around 10,830 Å. To target the 10,833-Å helium triplet, we binned the spectra from 10,590 to 11,150 Å into 20 narrowband channels. Each channel spanned four detector pixels, as a compromise between the low instrument resolution, the signal-to-noise ratio and the narrowness of the targeted feature. The wavelength coverage of each channel was shifted relative to the adjacent channel by one pixel, so the channels overlap.

We note that because the reported resolution¹⁰ of the G102 grism is $\lambda/\Delta\lambda \approx 155$ at 10,400 Å (which corresponds to $\Delta\lambda \approx 67$ Å, or 2.7 pixel widths), the smallest bins that are theoretically possible are three pixels wide. A resolution of three pixels could be achieved if the 10,833-Å feature lay in the centre of a pixel; however, in our data it lies considerably blue-wards of the centre of its pixel. This means there is some 10,833-Å flux in the pixel located two pixels blue-wards of the 10,833 Å line. Indeed, when we tested the three-pixel case, we found that the amplitude of the 10,833-Å feature increased by 0.011% compared with the four-pixel-bin fit, which is similar to the expected increase of 0.016% if all the 10,833-Å flux fell within a central three-pixel bin. With three-pixel bins the feature also appeared to have a slight blue ‘wing’, which is unlikely to have an astrophysical origin because such wings would be expected from binning the data to a resolution higher than that of the spectrograph. We therefore used conservative four-pixel bins.

Extended Data Fig. 2 shows the spectroscopic light curves divided by their best-fitting systematics models, along with their residuals. Extended Data Table 2 shows our median values for the transit depth and their 1σ uncertainties, calculated from the MCMC chains. We also list the root mean square of the residuals for each channel, which range from 0.976 to 1.22 relative to the photon noise. The resulting transmission spectrum is shown in Fig. 2.

Previous studies³¹ have highlighted the importance of considering the effect of stellar limb darkening in stellar absorption lines on exoplanet transmission spectra. To investigate whether this could cause the strong feature at 10,833 Å, we re-ran the narrowband spectroscopic light-curve fits while fitting for a quadratic limb-darkening law. The resulting spectrum was consistent with our previous analysis within 1σ .

Strong stellar lines that shift over the edges of pixels can introduce noise to measured transmission spectra³². We checked this effect by smoothing our extracted time-series spectra with a Gaussian kernel with a full-width at half-maximum of four pixels, and re-running the narrowband spectroscopic light-curve fits. Our measured 10,833-Å absorption feature remained consistent within 1σ .

MEarth observations. Photometric monitoring observations were performed using a single telescope of the MEarth–South^{33,34} array (CS 2015) at the Cerro Tololo Inter-American Observatory in Chile. Data were obtained on 78 nights from 22 March 2017 (universal time) to 1 August 2017 in groups of four 15-s exposures, with these exposure groups repeated at a cadence of approximately 30 min. A total of 3,096 exposures were collected over this period. The bandpass of these observations is in the red optical region, with the blue cut-off defined by RG715 glass to be approximately 7,150 Å and the red cut-off defined by the decline of the quantum efficiency of the charge-coupled device (CCD) to be approximately 10,000 Å. For our data reduction, we used our previously published methodology³⁵, modified for the specifics of the MEarth data³⁶. The CCD camera shutter failed on 9 May 2017, which required removal for servicing.

Because this procedure introduces flat-fielding errors that are not corrected to sufficient precision by standard calibrations, we allow for these errors explicitly in the analysis by solving for a change in the magnitude-zero points on both sides of the meridian at this date by following standard methods³⁷. The result of this analysis is a ‘least-squares periodogram’ (shown in Extended Data Fig. 3), obtained by simultaneously fitting a periodic modulation, while accounting for the four magnitude-zero points and two additional linear terms describing sources of systematic errors in the photometry (the full-width at half-maximum of the stellar images, and the ‘common mode’ as a proxy for the effect of variable precipitable water vapour on the photometry). This procedure would be mathematically equivalent to a Lomb–Scargle periodogram in the absence of these six extra terms. The

highest peak in the periodogram and its full-width at half-maximum correspond to a periodicity of 19.7 ± 0.9 days. This is consistent with estimates from Kepler light curves of 17.5 ± 1.4 days³⁰. We find an amplitude of about 0.00150 mag.

Automated imaging telescope photometry. We acquired nightly photometric observations of WASP-107 with the Tennessee State University Celestron 14-inch (C14) automated imaging telescope (AIT), which is located at the Fairborn Observatory in southern Arizona^{38,39}. The observations were made in the Cousins R passband with an SBIG STL-1001E CCD camera. Differential magnitudes of WASP-107 were computed with respect to eight of the most constant comparison stars in the CCD field. Details of our data acquisition, reduction and analysis can be found in a previous publication⁴⁰, which describes a similar analysis of the planetary host star WASP-31.

A total of 120 nightly observations (excluding a few observations in transit) were collected between 23 February and 28 June 2017. The nightly differential magnitudes are plotted in Extended Data Fig. 4a. Extended Data Fig. 4b and c shows the frequency spectrum of the observations and the phase curve computed with the best frequency. Our frequency analysis is based on least-squares sine fits with trial frequencies between 0.01 and 0.5 cycles per day, corresponding to periods between 2 and 100 days. The goodness of fit at each frequency is determined as the reduction factor in the variance of the original data. Low-amplitude brightness variability is seen at a period of 8.675 ± 0.043 days with a peak-to-peak amplitude of only 0.005 mag. This period is almost exactly half of the 17.5-day rotation period found from Kepler light curves³⁰ and demonstrates that WASP-107 has spots or spot groups on opposite hemispheres of the star during the epoch of our observations. The WASP-107b discovery team⁶ also found periods of around 17 and 8.3 days in their 2009 and 2010 photometry.

Stellar variability correction. To correct for stellar variability between the G141 and G102 epochs, we follow a method similar to that used in previous studies^{41,42}, and estimate the flux from the non-spotted stellar surface as $F_s = \max(F) + k\sigma$, where F is the photometric light curve, k is a fitted value and σ is the standard deviation of the light curve. We adopt the value $k = 1$, which was determined in a previous study⁴¹. We use the best-fitting period, amplitude and ephemeris from the MEarth photometry to estimate the expected flux-dimming correction at the mid-transit times for both datasets. We use the wavelength-dependent spot correction factor developed in a previous work⁴³ to correct for unocculted spots, and we set the spot temperature to be 3,200 K. After the correction, the two spectra align well and appear to share a flat baseline. The one overlapping spectral channel between G102 and G141 is consistent within 1σ .

ATMO retrieval. For the combined G102 and G141 broadband spectrum, corrected for photospheric variability, we performed an atmospheric retrieval analysis using our 1D radiative transfer code, ATMO^{12,13,44–46}. We assumed an isothermal temperature–pressure profile and used MCMC to fit for the following parameters: atmospheric temperature, planetary radius at a pressure of 1 mbar, grey cloud opacity, and the abundances of H₂O, CO₂, CO, CH₄, NH₃, H₂S, HCN and C₂H₂. We assumed solar abundances under chemical equilibrium for other gas species. We note that for this analysis we excluded wavelengths coinciding with the narrowband channel centred on the 10,833-Å helium triplet. Our best-fitting model is shown in Fig. 1, with $\chi^2 = 31.4$ for 18 degrees of freedom.

Assessing detector defects and random noise. We confirmed that the residuals for the pixel columns in each frame did not reveal any obvious anomalies over the narrow 10,833-Å helium triplet, which suggests that this feature is not caused by detector defects or uncorrected cosmic rays. In addition, the transit depths remained consistent within 0.5σ when we removed 1/3 of the points in the light curves, in several random sub-sets, and refitted them with the same procedures as described above.

Absorption from other species. The strong absorption line of metastable 2³S helium at 10,833 Å aligns extremely well with the peak of the feature. In the 20-Å region surrounding this peak (10,820 to 10,840 Å), helium is the only species that contains absorption solely within this wavelength range and nowhere else within the G102 bandpass (8,060 to 11,170 Å). There is, for example, a strong silicon absorption line at 10,830 Å and a water line at 10,835 Å (vacuum wavelengths)⁴⁷; however, if either species were the cause of the absorption seen in our transmission spectrum, there would be other similarly strong silicon lines measured at 10,588, 10,606 and 10,872 Å, and a water line at 10,929 Å, where we see no excess absorption. The other atoms with strong absorption lines near 10,833 Å are Np, Cs, Fe, Th, S, Cr, V, Yb, and Cu—all of which can be ruled out because they are either radioactive with short half-lives or have other strong transitions in the 8,060–11,170 Å wavelength range, which we do not observe. We have also found no species in the ExoMol⁴⁸ or HITRAN/HITEMP^{49,50} databases with sufficiently sharp features aligned at 10,833 Å. Specifically, we searched the following species: CH₄, CO₂, HCN, NH, CH, OH, PO, NO, VO, TiO, CN, C₂, PH₃, NH₃, SiO, CaO, H₃⁺, CO, H₂CO, C₂H₂, BeH, LiH, HCl, AlO, SO₂, H₂S, PN, KCl, NaCl, CS, CP, PS, MgH, NaH, CrH, CaH, FeH and ScH. We therefore conclude that absorption by metastable helium at 10,833 Å

is the most plausible explanation for the signal detected in the narrowband transmission spectrum.

Assessing Earth's exosphere. Where Earth's exosphere is illuminated by extreme ultraviolet radiation from the Sun, there is metastable helium. At an altitude of about 500 km, HST passes right through Earth's exosphere and, when it is not in Earth's shadow, it passes through regions containing metastable helium. The change in abundance of the metastable state throughout HST's orbit has been shown to impart a time-varying background signal in the 10,833-Å line on the timescale of one spacecraft orbit⁵¹ of about 95 min. There is no telluric metastable helium in Earth's shadow and, as expected, there is no substantial excess absorption at 10,833 Å while HST is in Earth's shadow⁵¹. It does, however, affect HST measurements at dawn and dusk—that is, when the spacecraft passes through the Sun-illuminated upper atmosphere. The magnitude of the effect is correlated with the solar activity cycle—that is, more activity means more ultraviolet radiation and more metastable helium. The effect of spatially diffuse telluric helium emission on WFC3 slit-less spectroscopy is to impart an increased sky background signal across the detector. At the time of the observations, we were approaching solar minimum, and the 10.7-cm radiation (which is a proxy for solar activity) was only 70 solar flux units (sfu), according to the Solar Monitoring Program of Natural Resources Canada (<http://www.spaceweather.gc.ca/solarflux/sx-en.php>). According to the WFC3 instrument report⁵¹, observations only appear considerably affected when the 10.7-cm flux is greater than about 100 sfu.

Nonetheless, to test whether metastable helium at dawn and dusk in Earth's atmosphere could cause an anomalous absorption feature in our transmission spectrum, we removed the first and last four exposures of each orbit—which encompass the initial and final 10 min—when HST passed through the illuminated dusk and dawn exosphere, and refitted the light curves. The results were consistent with previous analysis at less than 1σ , which indicates that emission from telluric helium is not the cause of the narrowband absorption feature in our data. We note that previous transit spectroscopic studies using G102^{52,53} did not show excess absorption at 10,833 Å.

Assessing the stellar chromosphere. We also considered the possibility that the absorption feature that we measured at 10,833 Å could be a result of stellar activity, as the metastable 2³S state of helium is formed in the inhomogeneous upper chromospheres and coronae of stars via photo-ionization, recombination and collisional excitation. The passage of the planet over quiet regions with less 10,833-Å helium absorption could, in theory, increase the relative transit depth at this wavelength and thus mimic an exoplanet atmospheric feature.

Theoretical models of chromospheres^{54,55} predict the maximum equivalent width of the 10,833-Å helium line in the spectra of F- to early K-type stars to be about 0.4 Å. Being a K6 star, WASP-107 lies just outside the valid range of spectral types for this model. However, in the following we show that to match our observed transmission spectral feature, the nominal chromospheric absorption of the WASP-107 host star at 10,833 Å would need to be five times stronger than any isolated (that is, non-multiple), main-sequence dwarf star measured until now.

After searching the literature for all 10,833-Å helium triplet equivalent-width measurements of isolated dwarf stars, we found more than 300 measurements of over 100 distinct stars, including 23 measurements of 11 different stars of similar spectral type to WASP-107 (K5–K7). We found no measurements greater than 0.409 Å^{56–61}. We took an additional measurement of the K6 star GJ380 with the Near-Infrared Echelle Spectrograph (NIRSpec) on Keck and found an equivalent width of 0.311 Å (A. Dupree, private communication).

Furthermore, it has been shown^{60,62} that the equivalent width of the 10,833-Å line is related to that of another neutral helium absorption line, at 5,876 Å. The 5,876-Å line is produced by the transition from the 2³D to the 2³P state. As such, the 5,876-Å line forms in the same regions of the stellar chromosphere as the 10,833-Å triplet (which corresponds to the 2³S–2³D transition). Extended Data Fig. 5 shows the equivalent-width measurements of the 10,833-Å and 5,876-Å lines in a survey of 31 FGK stars⁶⁰. A strong correlation is apparent.

To investigate the 5,876-Å helium line of WASP-107, we co-added high-resolution spectra obtained with the High Accuracy Radial velocity Planet Searcher (HARPS) spectrograph (European Southern Observatory programme 093.C-0474(A)). These spectra cover a wavelength range of 3,800–6,900 Å (Extended Data Fig. 5). We fitted the co-added spectrum for the equivalent width of the 5,876-Å helium line, with the result indicated in Extended Data Fig. 5 as a yellow-shaded region. We find that the equivalent width of this feature is similar to that measured for other single dwarf stars, with no evidence of unusual activity. Given the well established correlation between the equivalent widths of the 5,876-Å and 10,833-Å helium lines noted above, this provides further evidence against the WASP-107 host star having an abnormally deep 10,833-Å line. In addition, we measured the Ca II H- and K-line emission S-index (S_{HK}) for WASP-107 from the HARPS spectra and found a night-averaged value of $S_{HK} = 1.26 \pm 0.03$ (A.W., private communication), which is a moderate value for a K6 star⁶¹.

We therefore adopt the maximum equivalent width of 0.4 \AA to estimate an upper limit for the amplitude of a feature that could be caused by unocculted $10,833\text{-\AA}$ helium absorption of stellar origin in our 98-\AA -wide spectroscopic channel. We consider the limiting case in which WASP-107b occults only quiet regions of the star, where we assume there is no $10,833\text{-\AA}$ line absorption. This is the scenario in which the maximum amount of stellar-continuum flux at $10,833 \text{ \AA}$ would be blocked out by the planet, which we treat as a fully opaque disk. We estimate the increased transit depth to be

$$D_{\text{activity}} = \frac{A_{\text{pl}}}{1 - \frac{W_{\text{He}}}{W_{\text{bin}}}} = 2.064\% \pm 0.005\%$$

where $A_{\text{pl}} = 2.056\% \pm 0.005\%$ is the fraction of the stellar area occulted by the planet, $W_{\text{He}} = 0.4 \text{ \AA}$, is the maximum equivalent width of the stellar absorption feature, and W_{bin} is the width of the spectral bin (98 \AA). This gives an upper limit of the feature caused by stellar activity, $\delta D_{\text{activity}} = D_{\text{activity}} - A_{\text{pl}} = 0.008\% \pm 0.005\%$, which is less than one-fifth of the measured size of the feature ($0.049\% \pm 0.011\%$). We therefore conclude that the observed absorption feature cannot be caused by stellar chromospheric spatial inhomogeneity alone.

Resolution-linked bias. If an absorption line overlaps in both a stellar and a planetary atmosphere spectrum, and the line is unresolved in the measured transmission spectrum, then the planetary absorption can be underestimated. The effect is called resolution-linked bias (RLB)⁶³. For the $10,833\text{-\AA}$ line in the WASP-107 system this dilution effect competes with the possible overestimation of the signal from unocculted chromospherically active regions (as described in the Methods section ‘Assessing the stellar chromosphere’). The magnitudes of both effects depend on whether the planet transits in front of active or quiet regions of the star. The RLB effect would be largest if the planet transited only chromospherically active regions (which have the highest $10,833\text{-\AA}$ line absorption). We estimated the magnitude of the RLB effect in this limiting case following the method described in a previous publication⁶³ and assuming an equivalent width of 0.4 \AA for the $10,833\text{-\AA}$ stellar line. For a measured absorption excess of $0.049\% \pm 0.011\%$ in a 98-\AA bin centred on the $10,833\text{-\AA}$ line, we could be underestimating the planetary absorption by up to 0.009% (that is, about one-fifth of the measured signal). However, without knowing which part of the chromosphere the planet transits, the stellar line profile and the velocity structure of the planetary helium signature, we cannot accurately estimate the magnitudes of the competing effects.

Stellar flares. Because the $10,833\text{-\AA}$ He line appears in the emission of solar (and presumably stellar) flares⁶⁴, active stars like WASP-107 could show short-term variability in this line, which may be difficult to disentangle from a transiting planetary signal³¹. Flares are unlikely to mimic the signal that we detect perfectly, because the planet would need to pass in front of flaring regions of the star throughout the duration of the transit. Instead, unocculted flares could dilute atmospheric absorption of the $10,833\text{-\AA}$ He line. Visual inspection of the raw light curve of the spectroscopic channel centred on $10,833 \text{ \AA}$ shows no evidence of flare events. Additionally, the pre- and post-transit flux levels agree with each other, which would not be the case if there was substantial $10,833\text{-\AA}$ emission from the tail of a flare. As a precaution, we re-produced the narrowband transmission spectrum around the $10,833\text{-\AA}$ line using different combinations of the out-of-transit baseline: first, with only orbits 2 and 4; then, with orbits 1 and 3; and then with orbits 2 and 5. All three cases gave a $10,833\text{-\AA}$ absorption feature that were consistent within 1σ of our full fit.

Photospheric spots and faculae. To quantify the effect of a heterogeneous photosphere on the transmission spectrum around $10,833 \text{ \AA}$, we used a variability modelling method^{65,66} that uses an ensemble of model stellar photospheres with randomly located active regions to estimate the fraction of the stellar surface covered by photospheric spots and faculae for a given rotational variability amplitude. While variability monitoring traces only the non-axisymmetric component of the stellar heterogeneity, and thus provides a lower limit on active-region-covering fractions⁶⁶, this numerical approach provides a more complete understanding of the range of covering fractions that may correspond to an observed variability level. The model describes the integrated full-disk spectrum by the combination of three components: the immaculate photosphere, spots and faculae. We used three spectra interpolated from the PHOENIX model grid⁶⁷ with $\log(g) = 4.5$, metallicity $[M/H] = +0.02$ and different temperatures to represent the three components. Following previous works⁶⁶, we set the photosphere temperature, T_{phot} , as the effective temperature of the star ($T_{\text{eff}} = 4,430 \text{ K}$)⁶ and adopted scaling relations for the spot temperature T_{spot} ^{68,69} and faculae temperature T_{fac} ⁷⁰.

Thus, the temperatures of the three components are $T_{\text{phot}} = T_{\text{eff}} = 4,430 \text{ K}$, $T_{\text{spot}} = 0.73 \times T_{\text{phot}} = 3,230 \text{ K}$, and $T_{\text{fac}} = T_{\text{phot}} + 100 \text{ K} = 4,530 \text{ K}$. The paper⁶ describing the discovery of WASP-107b reports a 17-day periodic modulation with a 0.4% semi-amplitude (0.8% full amplitude) for WASP-107. Assuming a typical spot radius of $r_{\text{spot}} = 2^\circ$, we find that the reported rotational variability could be caused by a spot filling fraction of $f_{\text{spot}} = (4^{+9}_{-2})\%$ (1σ confidence interval) if the

variability is due to spots alone. In the more realistic case, in which spots and faculae are both contributing to the variability, we find $f_{\text{spot}} = (8^{+6}_{-3})\%$ and $f_{\text{fac}} = (53^{+15}_{-12})\%$. The covering fractions that we report are means over the entire model photosphere; they do not take into account relative over- or under-abundances of magnetic features on the Earth-facing hemisphere during a transit. Therefore, in the worst-case scenario, they could underestimate the hemispheric covering fractions by a factor of 2. However, the 1σ confidence intervals, which are derived from 100 model realizations with randomly selected active region locations, are deliberately conservative to account for this. Extended Data Fig. 6 shows how unocculted photospheric stellar heterogeneities could affect the transmission spectrum, assuming that the planet transits a chord of immaculate photosphere. The stellar contamination factor, ϵ , on the y axis is multiplied by the true $(R_p/R_s)^2$ transit depth to produce the observed transmission spectrum; that is, $\epsilon > 1$ means that the observed transit depth is larger than expected from the planetary atmosphere model. The spots-plus-faculae model does not predict an increase in transit depth at $10,833 \text{ \AA}$. No sharp features are apparent around $10,833 \text{ \AA}$. Instead, the model predicts that transit depths should be inflated by about 1% across the full wavelength range of G102, with perhaps some features apparent at about $8,500 \text{ \AA}$ and $8,900 \text{ \AA}$ (for this reason, we only use the $8,780\text{--}11,370 \text{ \AA}$ region in our full transmission spectrum, even though the G102 throughput extends down to $8,000 \text{ \AA}$). The strong absorption feature that we measure is therefore unlikely to be caused by photospheric inhomogeneity.

1D escaping-atmosphere model. Here we give a brief overview of the first model that we used to investigate the narrowband transmission spectrum at $10,833 \text{ \AA}$, which is described in more detail in a previous publication¹⁴. This 1D model is based on the assumption that a thermosphere of a close-in exoplanet can be well represented by the density and velocity profiles of an isothermal Parker wind driven by gas pressure⁷¹. We assume a composition of atomic hydrogen (90% by number) and helium (10%). We find the solution for the hydrogen ionization balance and the distribution of helium atoms in the ground, excited 2^3S and ionized states. The physical processes taken into account in the helium balance are photoionization from the ground and 2^3S states, recombination to the singlet and triplet states, collisional transitions between the triplet 2^3S state and states in the helium singlet ladder (which includes collisions with both free electrons and neutral hydrogen atoms) and the radiative decay from the 2^3S state to the ground state. The photoionization rates are calculated using the ultraviolet stellar flux of a K6 star, HD 85512, taken from the MUSCLES survey⁷² (version 2.1^{73,74}), placed at the orbital distance of WASP-107b. The equations used to compute the hydrogen and helium distributions, along with all the relevant reaction rate coefficients and cross-sections, are described in a previous paper¹⁴. We changed only the input parameters—such as the mass and radius of the planet and its host star, as well as the input stellar spectrum—so that they match the properties of WASP-107b.

On the basis of the obtained density profile of helium in the 2^3S state, we calculate the optical depth and the in-transit absorption signal at $10,833 \text{ \AA}$, assuming that a planet with a spherically symmetric thermosphere transits across the centre of the stellar disk. For a planet of given mass and radius, the wind temperature and the total mass-loss rate are free parameters in the model. Using results from the literature^{75,76}, we explore the temperature range $5,000\text{--}13,000 \text{ K}$. To produce an absorption signal consistent with our measurement, the required mass-loss rate is between 10^{10} g s^{-1} and $3 \times 10^{11} \text{ g s}^{-1}$.

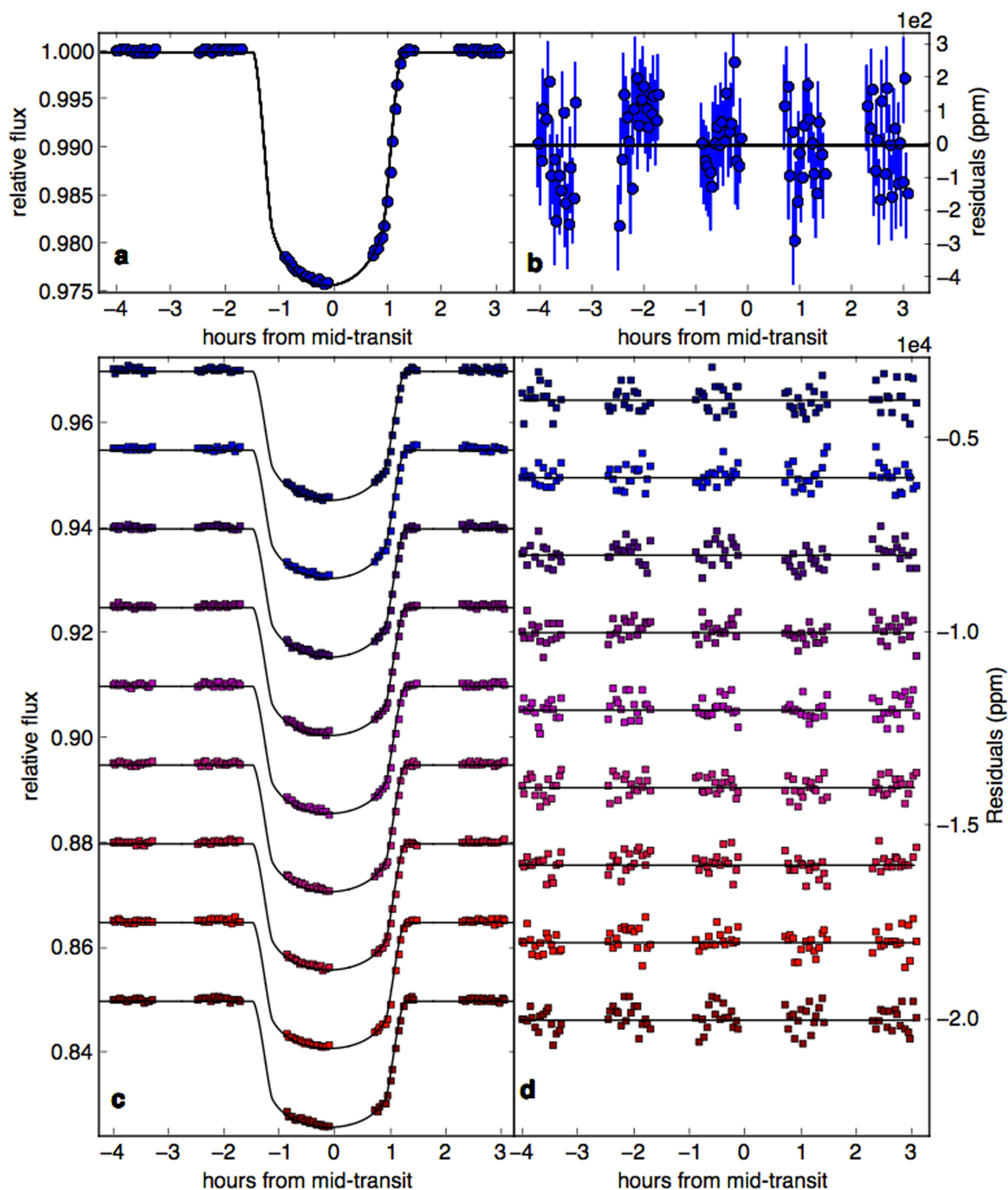
3D escaping-atmosphere model. Our second model has previously been used to interpret the escaping exosphere of the Neptune-mass exoplanet GJ436b^{16,77}. The model considers neutral helium atoms that are released from the top of the thermosphere and subjected to planetary and stellar gravity, radiation pressure and photoionization. We found that the data are well explained by 2^3S helium atoms escaping at a rate of $10^6\text{--}10^7 \text{ g s}^{-1}$. The stellar radiation pressure on the escaping helium atoms is stronger than the counter-balancing stellar gravity by a factor of approximately 10 and 50 for the weakest and strongest of the $10,833\text{-\AA}$ triplet lines, respectively. Thus, the gas is blown away from the exoplanet so swiftly as to form a tail nearly aligned with the star-planet axis.

Code availability. The custom code used to extract the HST spectra from the raw data frames is available upon request. The HST light-curve fitting was performed using the open-source BATMAN (<https://github.com/lkreidberg/batman>) and emcee codes (<http://github.com/dfm/emcee>), and the proprietary RECTE code. The ATMO code used to compute the lower-atmosphere models is currently proprietary, as are the 1D and 3D upper-atmosphere codes.

Data availability. Raw HST data frames are publicly available online at the Mikulski Archive for Space Telescopes (MAST; <https://archive.stsci.edu>).

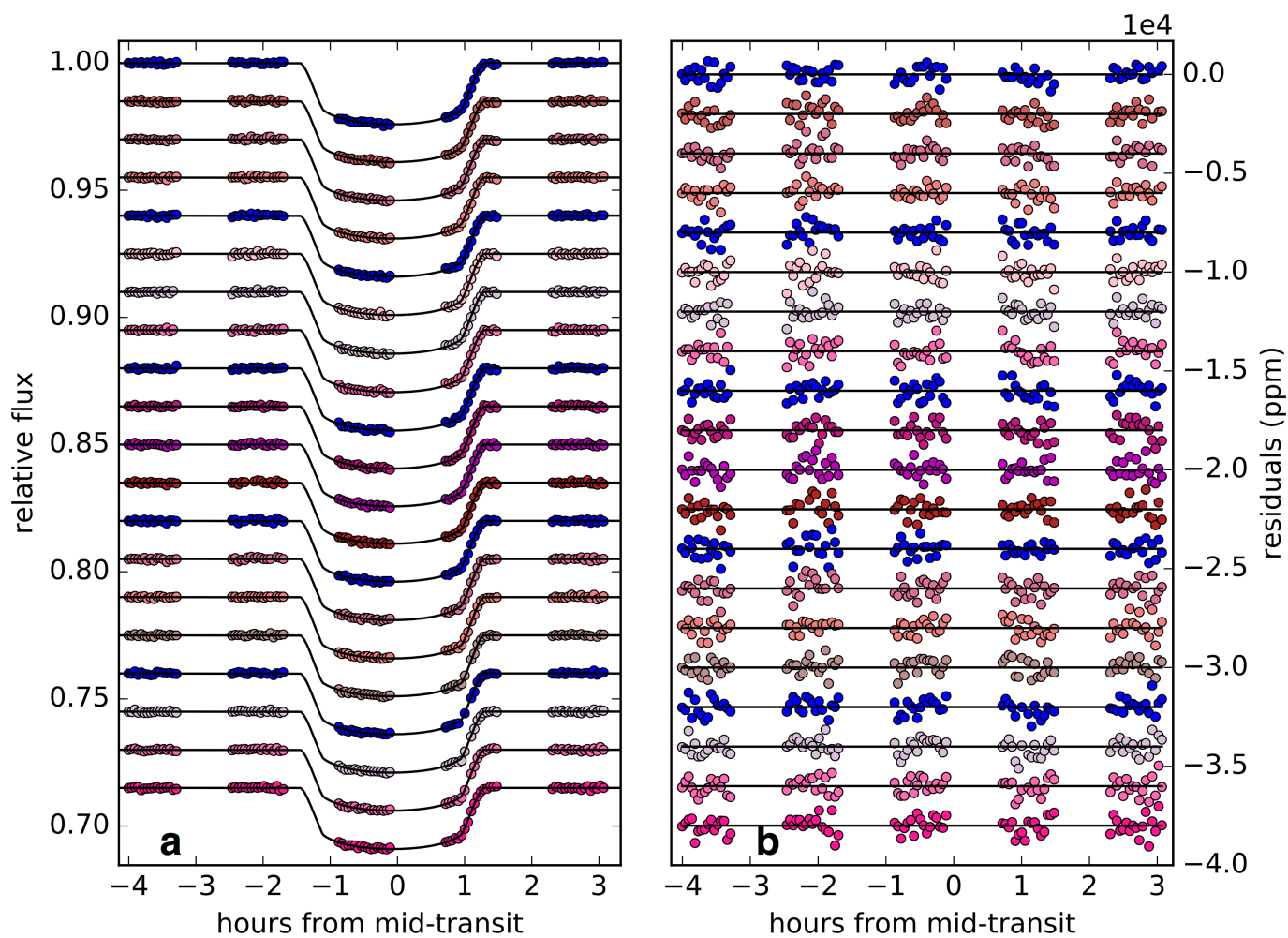
26. Ryan, R. E. Jr et al. *The Updated Calibration Pipeline for WFC3/UVIS: A Reference Guide to CALWF3 (Version 3.3)*. Report No. WFC3-2016-001 (Space Telescope Science Institute, 2016); <http://www.stsci.edu/hst/wfc3/documents/ISRs/WFC3-2016-01.pdf>.
27. Evans, T. M. et al. Detection of H_2O and evidence for TiO/VO in an ultra-hot exoplanet atmosphere. *Astrophys. J.* **822**, L4 (2016).

28. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for $-5.0 \leq \log[M/H] \leq +1$, $2000 \text{ K} \leq T_{\text{eff}} \leq 50000 \text{ K}$ at several surface gravities. *Astron. Astrophys.* **363**, 1081–1190 (2000).
29. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
30. Močnik, T., Hellier, C., Anderson, D. R., Clark, B. J. M. & Southworth, J. Starspots on WASP-107 and pulsations of WASP-118. *Mon. Not. R. Astron. Soc.* **469**, 1622–1629 (2017).
31. Czesla, S., Klocová, T., Khalafinejad, S., Wolter, U. & Schmitt, J. H. M. M. The center-to-limb variation across the Fraunhofer lines of HD 189733. Sampling the stellar spectrum using a transiting planet. *Astron. Astrophys.* **582**, A51 (2015).
32. Deming, D. et al. Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophys. J.* **774**, 95 (2013).
33. Nutzman, P. & Charbonneau, D. Design considerations for a ground-based transit search for habitable planets orbiting M dwarfs. *Publ. Astron. Soc. Pacif.* **120**, 317 (2008).
34. Irwin, J. M. et al. The MEarth-North and MEarth-South transit surveys: searching for habitable super-Earth exoplanets around nearby M-dwarfs. In *Proc. 18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun* (eds van Belle, G. & Harris, H. C.) 767–762 (2014); preprint at <https://arxiv.org/abs/1409.0891>.
35. Irwin, J. et al. The Monitor project: rotation of low-mass stars in the open cluster M34. *Mon. Not. R. Astron. Soc.* **370**, 954 (2006).
36. Berta, Z. K. et al. Transit detection in the MEarth survey of nearby M dwarfs: bridging the clean-first, search-later divide. *Astron. J.* **144**, 145 (2012).
37. Irwin, J. et al. On the angular momentum evolution of fully convective stars: rotation periods for field M-dwarfs from the MEarth transit survey. *Astrophys. J.* **727**, 56 (2011).
38. Henry, G. W. Techniques for automated high-precision photometry of Sun-like stars. *Publ. Astron. Soc. Pacif.* **111**, 845–860 (1999).
39. Eaton, J. A., Henry, G. W. & Fekel, F. C. in *The Future of Small Telescopes in the New Millennium, Volume II – The Telescopes We Use* (ed. Oswalt T. D.) 189–207 (Kluwer, Dordrecht, 2003).
40. Sing, D. K. et al. HST hot-Jupiter transmission spectral survey: detection of potassium in WASP-31b along with a cloud deck and Rayleigh scattering. *Mon. Not. R. Astron. Soc.* **446**, 2428–2443 (2015).
41. Aigrain, S., Pont, F. & Zucker, S. A simple method to estimate radial velocity variations due to stellar activity using photometry. *Mon. Not. R. Astron. Soc.* **419**, 3147–3158 (2012).
42. Huitson, C. et al. An HST optical-to-near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO. *Mon. Not. R. Astron. Soc.* **434**, 3252–3274 (2013).
43. Sing, D. K. et al. Hubble Space Telescope transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *Mon. Not. R. Astron. Soc.* **416**, 1443–1455 (2011).
44. Tremblin, P. et al. Cloudless Atmospheres for L/T Dwarfs and Extrasolar Giant Planets. *Astrophys. J.* **817**, L19 (2016).
45. Wakeford, H. R. et al. HAT-P-26b: A Neptune-mass exoplanet with a well-constrained heavy element abundance. *Science* **356**, 628–631 (2017).
46. Evans, T. M. et al. An ultrahot gas-giant exoplanet with a stratosphere. *Nature* **548**, 58–61 (2017).
47. Kramida, A. et al. *NIST Atomic Spectra Database (version 5.5.1)* <https://physics.nist.gov/asd> (2017).
48. Tennyson, J. et al. The ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J. Mol. Spectrosc.* **327**, 73–94 (2016).
49. Gordon, I. E. et al. The HITRAN 2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **203**, 3–69 (2017).
50. Rothman, L. S. et al. HITEMP, the high-temperature molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **111**, 2139–2150 (2010).
51. Brammer, G., Pirzkal, N., McCullough, P. & MacKenty, J. *Time-varying Excess Earth-glow Backgrounds in the WFC3/IR Channel*. Report No. WFC3 2014-03 (Space Telescope Science Institute, 2014); <http://www.stsci.edu/hst/wfc3/documents/ISRs/WFC3-2014-03.pdf>.
52. Kreidberg, L. et al. A detection of water in the transmission spectrum of the hot Jupiter WASP-12b and implications for its atmospheric composition. *Astrophys. J.* **814**, 66 (2015).
53. Wakeford, H. et al. The complete transmission spectrum of WASP-39b with a precise water constraint. *Astron. J.* **155**, 29 (2017).
54. Andretta, V. & Jones, H. P. On the role of the solar corona and transition region in the excitation of the spectrum of neutral helium. *Astrophys. J.* **489**, 375–394 (1997).
55. Vaughan, A. H. & Zirin, H. The helium line $\lambda 10830 \text{ \AA}$ in late-type stars. *Astrophys. J.* **152**, 123–139 (1968).
56. Zirin, H. $\lambda 10830 \text{ He I}$ observations of 455 stars. *Astrophys. J.* **260**, 655–669 (1982).
57. Zarro, D. M. & Zirin, H. The dependence of He I $\lambda 10830$ absorption strength upon X-ray emission in late-type stars. *Astrophys. J.* **304**, 365–370 (1986).
58. Sanz-Forcada, J. & Dupree, A. K. Active cool stars and He I 10830 \AA : the coronal connection. *Astron. Astrophys.* **488**, 715–721 (2008).
59. Takeda, Y. & Takada-Hidai, M. Chromospheres in metal-poor stars evidenced from the He I 10830 \AA line. *Publ. Astron. Soc. Jpn.* **63**, S547–S554 (2011).
60. Andretta, V., Giampapa, M. S., Covino, E., Reiners, A. & Beeck, B. Estimates of active region area coverage through simultaneous measurements of the He I $\lambda \lambda 5876$ and 10830 lines. *Astrophys. J.* **839**, 97 (2017).
61. Isaacson, H. I. & Fischer, D. Chromospheric activity and jitter measurements for 2630 stars on the California planet search. *Astrophys. J.* **725**, 875–885 (2010).
62. Andretta, V. & Giampapa, M. S. A method for estimating the fractional area coverage of active regions on dwarf F and G stars. *Astrophys. J.* **439**, 405–416 (1995).
63. Deming, D. & Sheppard, K. Spectral resolution-linked bias in transit spectroscopy of extrasolar planets. *Astrophys. J.* **841**, L3 (2017).
64. Li, H., You, J., Yu, X. & Du, Q. Spectral characteristics of solar flares in different chromospheric lines and their implications. *Sol. Phys.* **241**, 301–315 (2007).
65. Rackham, B. et al. ACCESS I: an optical transmission spectrum of GJ 1214b reveals a heterogeneous stellar photosphere. *Astrophys. J.* **834**, 151 (2017).
66. Rackham, B. V., Apai, D. & Giampapa, M. S. The transit light source effect: false spectral features and incorrect densities for M-dwarf transiting planets. *Astrophys. J.* **853**, 122 (2018).
67. Husser, T.-O. et al. A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. *Astron. Astrophys.* **553**, A6 (2013).
68. Beryugina, S. Starspots: a key to the stellar dynamo. *Living Rev. Sol. Phys.* **2**, 8 (2005).
69. Afram, N. & Beryudina, S. Molecules as magnetic probes of starspots. *Astron. Astrophys.* **576**, A34 (2015).
70. Gondoin, P. Contribution of Sun-like faculae to the light-curve modulation of young active dwarfs. *Astron. Astrophys.* **478**, 883–887 (2008).
71. Parker, E. N. Dynamics of the interplanetary gas and magnetic fields. *Astrophys. J.* **128**, 664–676 (1958).
72. France, K. et al. The MUSCLES treasury survey. I. Motivation and overview. *Astrophys. J.* **820**, 89 (2016).
73. Youngblood, A. et al. The MUSCLES treasury survey. II. Intrinsic $\text{Ly}\alpha$ and extreme ultraviolet spectra of K and M dwarfs with exoplanets. *Astrophys. J.* **824**, 101 (2016).
74. Loyd, R. O. P. et al. The MUSCLES treasury survey. III. X-ray to infrared spectra of 11 M and K stars hosting planets. *Astrophys. J.* **824**, 102 (2016).
75. Lecavelier des Etangs, A., Vidal-Madjar, A., McConnell, J. C. & Hebrard, G. Atmospheric escape from hot Jupiters. *Astron. Astrophys.* **418**, L1–L4 (2004).
76. Salz, M., Czesla, S., Schneider, P. C. & Schmitt, J. H. M. M. Simulating the escaping atmospheres of hot gas planets in the solar neighborhood. *Astron. Astrophys.* **586**, A75 (2016).
77. Ehrenreich, D. et al. A giant comet-like cloud of hydrogen escaping the warm Neptune-mass exoplanet GJ 436b. *Nature* **522**, 459–461 (2015).



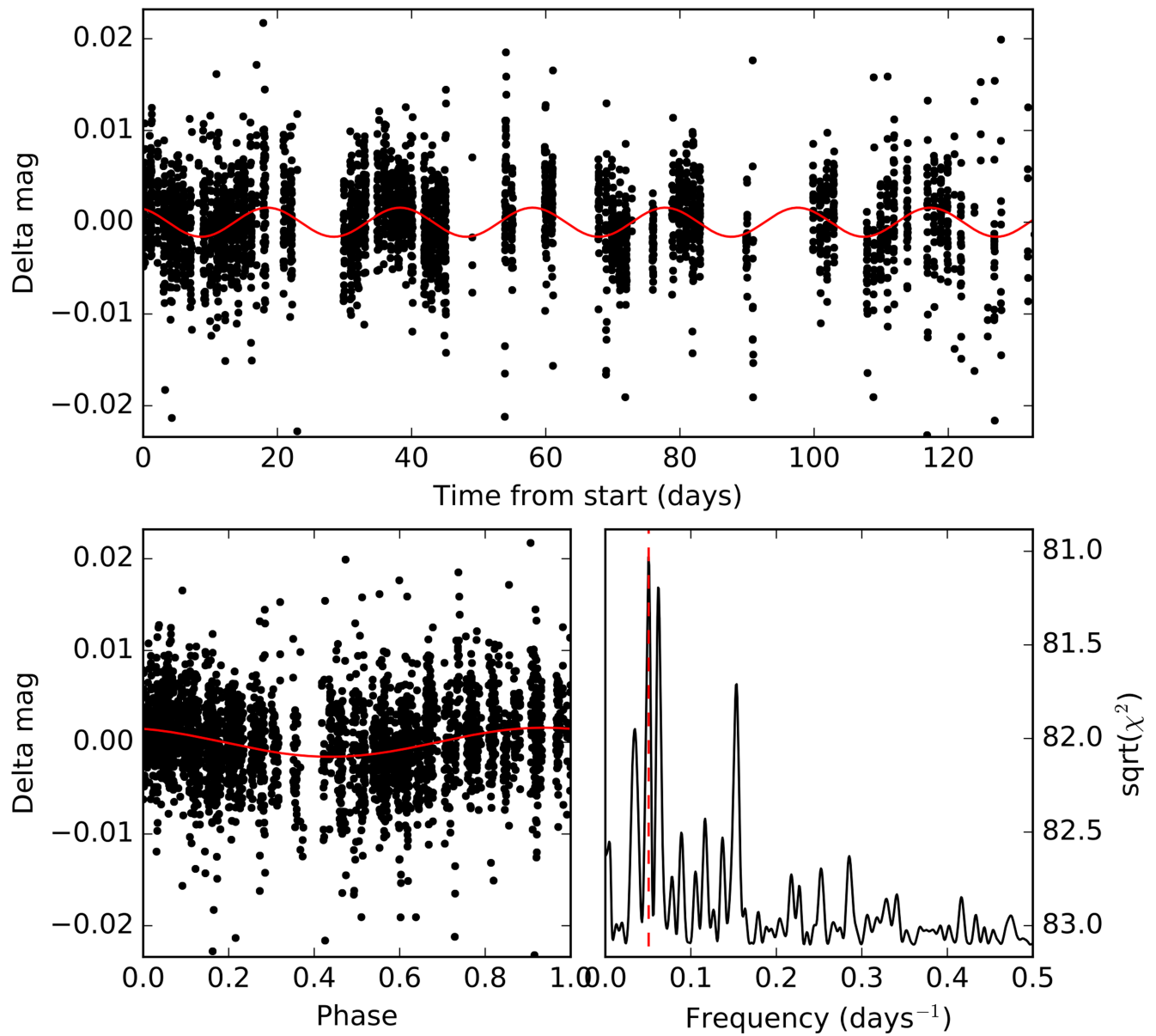
Extended Data Fig. 1 | G102 white-light curve and broadband spectroscopic light curves covering the wavelength range 0.88–1.14 μm for WASP-107b. **a, Relative flux of the white-light curve with respect to systematics model results (blue points), with the best-fitting transit light curve plotted in black. **b**, White-light residuals and 1σ errors, after removing the combined transit and systematics components of the**

best-fitting model. **c**, Points are spectroscopic light curves divided by systematics model results, and black curves are best-fitting transit models, with vertical offsets applied for clarity. **d**, Best-fitting spectroscopic model residuals, with vertical offsets applied for clarity. Differently coloured points in **c** and **d** are used to highlight separate channels.



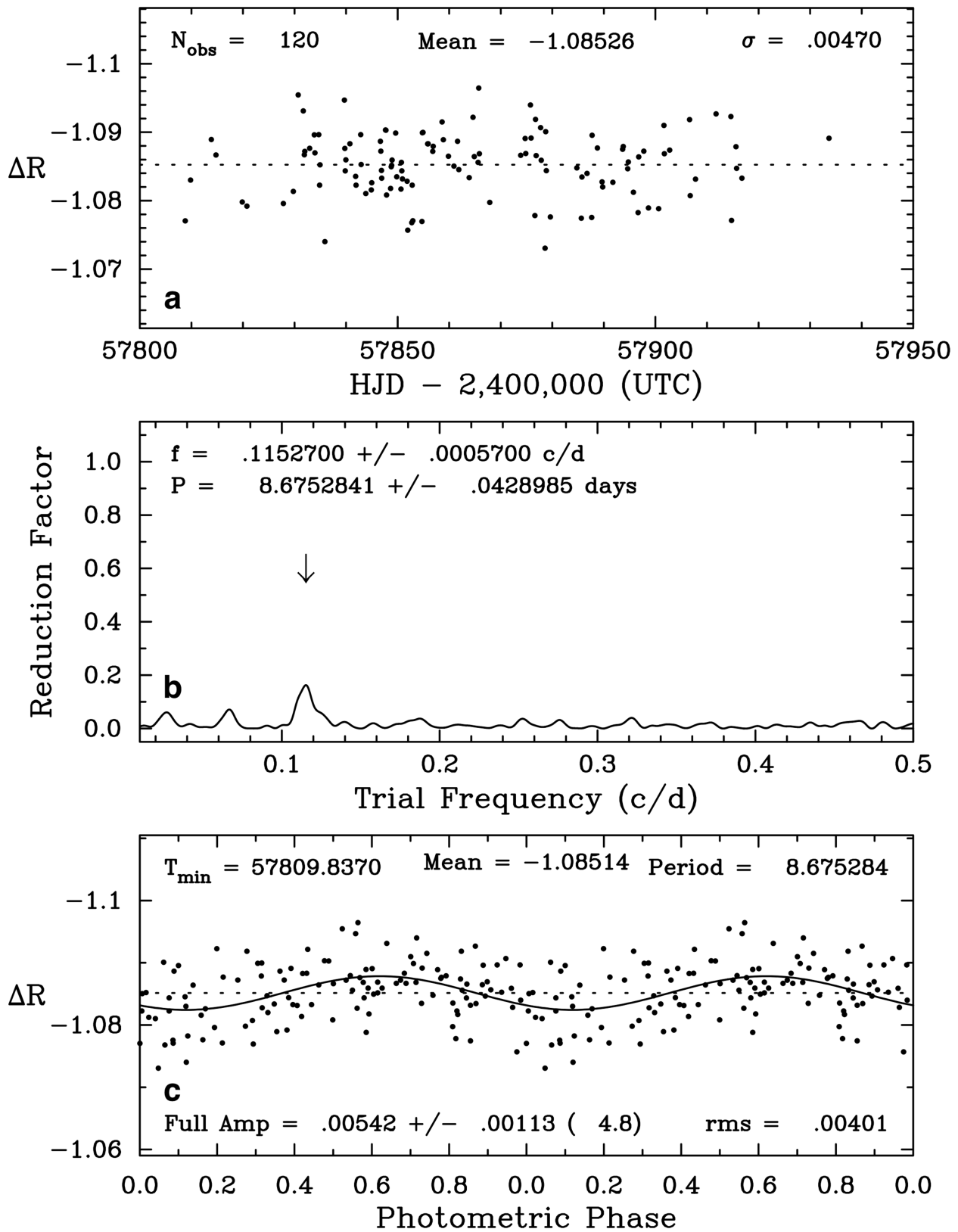
Extended Data Fig. 2 | Narrowband (four-pixel-wide) spectroscopic light curves covering the wavelength range 1.06–1.12 μm . **a**, Points are light curves divided by systematics model results, black curves are best-fitting transit models. **b**, Best-fitting model residuals, with vertical

offsets applied for clarity. Differently coloured points correspond to different channels. The five non-overlapping channels used to measure the 10,833-Å line absorption are highlighted in blue.



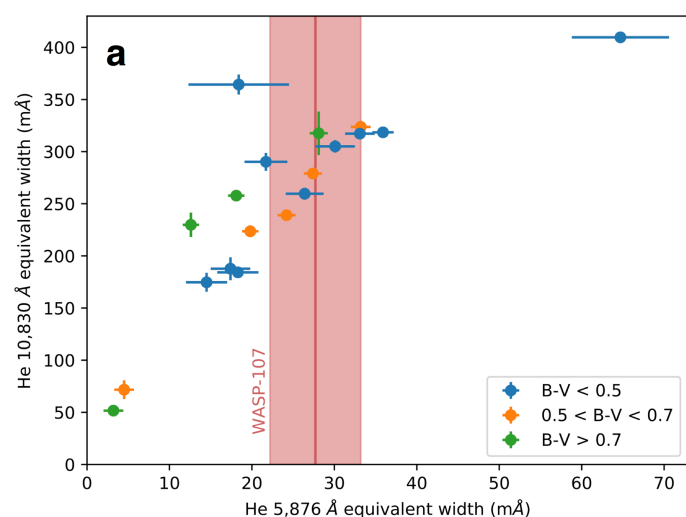
Extended Data Fig. 3 | Ground-based photometry for WASP-107 from MEarth. We performed a Lomb–Scargle periodogram search and found a best-fitting period of 19.7 ± 0.9 days (dashed red line), with a relative

amplitude of about 0.15%. Solid red lines show the results of the best-fitting sinusoidal model.

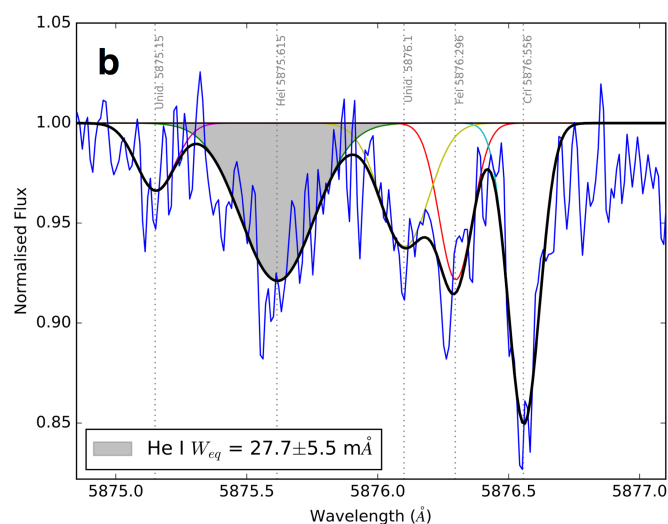


Extended Data Fig. 4 | Ground-based photometry for WASP-107b from AIT. **a**, Nightly photometric observations of WASP-107, acquired with Tennessee State University's C14 AIT at the Fairborn Observatory during the 2017 observing season. The number of observations (N_{obs}) was 120. ΔR is the relative flux in the Cousins R band. **b**, The frequency spectrum

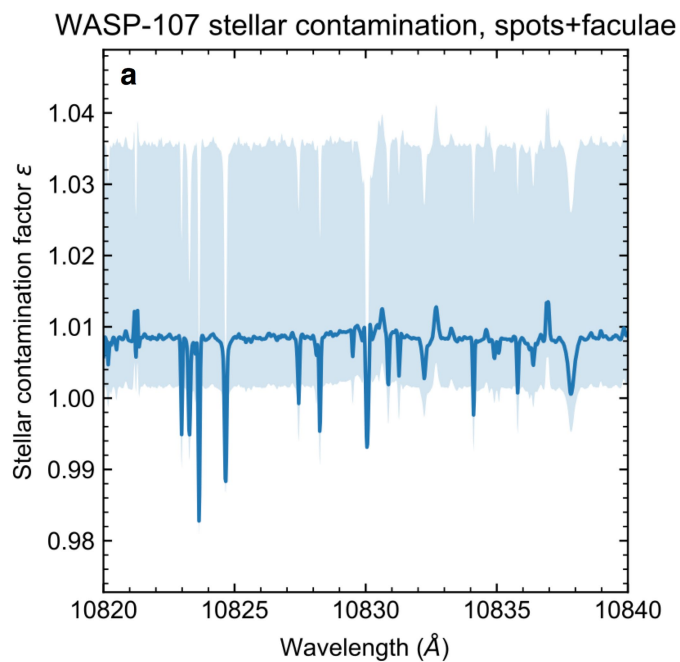
of the 2017 observations shows low-amplitude variability with a period (P) of 8.675 days (a frequency, f , of 0.115 cycles per day). **c**, The data phased to the 8.675-day period have a peak-to-peak amplitude of only 0.005 mag. HJD, heliocentric Julian Date; UCT, coordinated universal time; c/d, cycles per day; T_{min} , best-fit ephemeris.



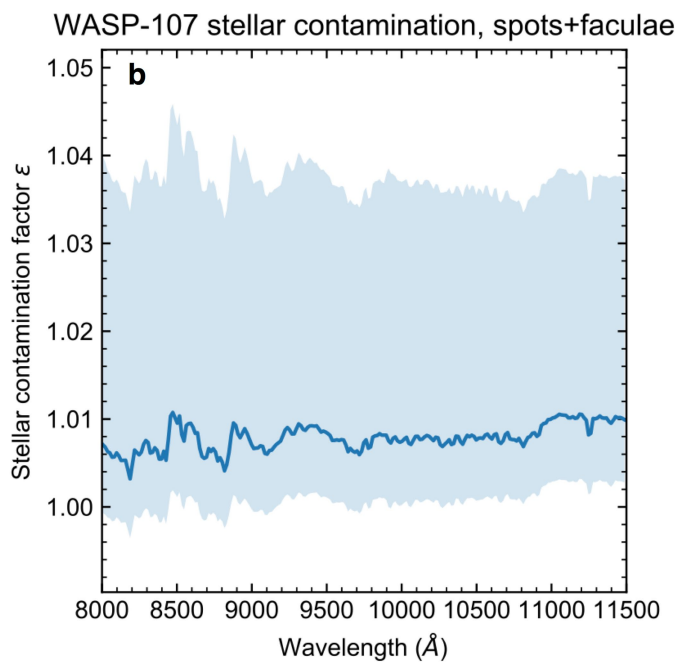
Extended Data Fig. 5 | Equivalent widths of helium 5,876-Å and 10,830-Å lines. **a**, Measurements for 30 stars of different colour indices with 1σ errors from a previous work⁶⁰. The two helium lines are expected to form in the same stellar-atmosphere regions and their equivalent widths are clearly correlated. Our 5,876-Å line measurement for WASP-107 (colour index $B-V > 0.7$), obtained from HARPS spectra, is plotted



as a red line and the red-shaded region shows the 1σ error. **b**, Co-added spectra of WASP-107b around the 5,876-Å line of metastable helium from the HARPS radial-velocity campaign (blue line). Absorption lines are fitted with Gaussian profiles, and the best-fit results are shown as green, yellow and red lines, with the sum of the profiles shown in black. The best-fitting line profile of the 5,876-Å line is shaded in grey.



Extended Data Fig. 6 | Effects of an inhomogeneous photosphere on the transmission spectrum of WASP-107b. Lines show the stellar contamination produced by unocculted spots and faculae. Shaded regions indicate the 1σ uncertainty on the stellar contamination due to



the uncertainty on spot- and faculae-covering fractions. **a**, The region around the 10,830- \AA (air wavelength) helium triplet at the resolution of the PHOENIX spectra (resolving power, $R = 500,000$). **(b)** The full G102 wavelength range in 15- \AA bins.

Extended Data Table 1 | Fitted parameters from the G102 white-light curve

Parameter	Value
R_P/R_S	0.142988 ± 0.00012
t_0 (BJD _{UTC})	$2,457,904.7295 \pm 0.0002$
c_0	$1.00004 \pm 2 \times 10^{-5}$
c_1	-0.0018 ± 0.0002
s_{pop}	62 ± 17
f_{pop}	42 ± 6
δs	-2 ± 10
δf	65 ± 4
β	1.73 ± 0.15
P	5.72147^a
$i(^{\circ})$	89.7^a
a/R_S	18.164^a
e (assumed)	0

Errors quoted were calculated using 68% of the MCMC samples after burn-in. BJD, barycentric Julian date; UTC, coordinated universal time.

^aParameters fixed from Dai & Winn⁷.

Extended Data Table 2 | All results from transit light-curve fits

Wavelength (Å)	Transit depth (%)	Error (%)	RMS (PPM)	RMS/ phot.	Correction factor
8,769 - 9,063	2.0451	0.0084	326	1.178	1.007101
9,063 - 9,356	2.0425	0.0069	276	1.077	1.006785
9,356 - 9,650	2.0514	0.0079	285	1.184	1.006549
9,650 - 9,943	2.0514	0.0064	252	1.083	1.006454
9,943 - 10,237	2.0456	0.0066	264	1.167	1.006340
10,237 - 10,530	2.0448	0.0058	241	1.080	1.006303
10,530 - 10,775	2.0431	0.0065	245	1.048	1.006162
10,775 - 11,142	2.0461	0.007	269	1.152	1.006123
11,142 - 11,386	2.0509	0.0069	298	1.198	1.005945
10,579 - 10,677	2.0634	0.0091	344	0.989	1.00596
10,604 - 10,701	2.0500	0.0088	381	1.102	1.005923
10,628 - 10,726	2.0604	0.0089	366	1.061	1.006214
10,652 - 10,750	2.0571	0.0075	336	0.976	1.006167
10,677 - 10,775	2.0563	0.0082	360	1.043	1.006131
10,701 - 10,799	2.0643	0.0103	395	1.143	1.006046
10,726 - 10,824	2.0830	0.0094	354	1.023	1.005985
10,750 - 10,848	2.0964	0.0102	415	1.198	1.005928
10,775 - 10,873	2.1048	0.0097	391	1.126	1.005923
10,799 - 10,897	2.0998	0.0084	387	1.117	1.005948
10,824 - 10,922	2.0870	0.0091	390	1.128	1.005949
10,848 - 10,946	2.0585	0.0095	409	1.183	1.006008
10,873 - 10,970	2.0546	0.0104	385	1.111	1.005982
10,897 - 10,995	2.0634	0.0108	423	1.220	1.005973
10,922 - 11,019	2.0642	0.0098	377	1.087	1.005967
10,946 - 11,044	2.0543	0.0093	363	1.046	1.005935
10,970 - 11,068	2.0502	0.0101	375	1.084	1.005962
10,995 - 11,093	2.0584	0.0103	373	1.082	1.005918
11,019 - 11,117	2.0564	0.0098	385	1.117	1.005897
11,044 - 11,142	2.0631	0.0105	414	1.197	1.005891
Modified, previously published results ¹⁸ :					
11,210 - 11,450	2.0723	0.0059			1.003979
11,450 - 11,710	2.0814	0.0055			1.003919
11,710 - 11,960	2.0585	0.0056			1.003918
11,960 - 12,220	2.0577	0.0054			1.003848
12,220 - 12,480	2.0535	0.0059			1.003892
12,480 - 12,720	2.0572	0.0050			1.003897
12,720 - 12,980	2.0699	0.0062			1.003830
12,980 - 13,230	2.0818	0.0050			1.003805
13,230 - 13,490	2.0742	0.0057			1.003983
13,490 - 13,740	2.0943	0.0048			1.004081
13,740 - 14,010	2.0878	0.0048			1.004059
14,010 - 14,250	2.0974	0.0052			1.004110
14,250 - 14,520	2.0907	0.0062			1.004126
14,520 - 14,760	2.0777	0.0051			1.004136
14,760 - 15,020	2.0767	0.0069			1.004107
15,020 - 15,280	2.0762	0.0067			1.004020
15,280 - 15,520	2.0593	0.0060			1.004116
15,520 - 15,790	2.0562	0.0064			1.004007
15,790 - 16,030	2.0581	0.0056			1.003941
16,030 - 16,290	2.0595	0.0065			1.003969

Modified results from a previous study³ are included. RMS denotes the root mean square of the model residuals in parts per million (PPM). The second-to-last column is the RMS divided by the expected photon noise. The last column is the correction factor that we applied to account for stellar variability.

Extended Data Table 3 | Results from ATMO retrieval code for the lower atmosphere

Parameter	Limits from MCMC
Temperature (K)	650^{+120}_{-80}
R_p/R_s at 1mbar	$0.914^{+0.010}_{-0.014}$
VMR $\log_{10}(\text{H}_2\text{O})$	$-1.7^{+0.3}_{-0.6}$
VMR $\log_{10}(\text{CO}_2)$	<10
VMR $\log_{10}(\text{CO})$	<11
VMR $\log_{10}(\text{CH}_4)$	<10
VMR $\log_{10}(\text{NH}_3)$	<10
VMR $\log_{10}(\text{H}_2\text{S})$	<11
VMR $\log_{10}(\text{HCN})$	<11
VMR $\log_{10}(\text{C}_2\text{H}_2)$	<10

VMR stands for volume mixing ratio. Uncertainties for the temperature, R_p/R_s and the VMR of H_2O were calculated using 68% of the MCMC samples after burn-in. Upper limits are from 1σ MCMC errors.

Characterization of the 1S–2S transition in antihydrogen

M. Ahmadi¹, B. X. R. Alves², C. J. Baker³, W. Bertsche^{4,5}, A. Capra⁶, C. Carruth⁷, C. L. Cesar⁸, M. Charlton³, S. Cohen⁹, R. Collister⁶, S. Eriksson³, A. Evans¹⁰, N. Evetts¹¹, J. Fajans⁷, T. Friesen², M. C. Fujiwara⁶, D. R. Gill⁶, J. S. Hangst^{2*}, W. N. Hardy¹¹, M. E. Hayden¹², C. A. Isaac³, M. A. Johnson^{4,5}, J. M. Jones³, S. A. Jones^{2,3}, S. Jonsell¹³, A. Khramov⁶, P. Knapp³, L. Kurchaninov⁶, N. Madsen³, D. Maxwell³, J. T. K. McKenna⁶, S. Menary¹⁴, T. Momose¹¹, J. J. Munich¹², K. Olchanski⁶, A. Olin^{6,15}, P. Pusa¹, C. Ø. Rasmussen², F. Robicheaux¹⁶, R. L. Sacramento⁸, M. Sameed^{3,4}, E. Sarid¹⁷, D. M. Silveira⁸, G. Stutter², C. So¹⁰, T. D. Tharp¹⁸, R. I. Thompson¹⁰, D. P. van der Werf^{3,19} & J. S. Wurtele⁷

In 1928, Dirac published an equation¹ that combined quantum mechanics and special relativity. Negative-energy solutions to this equation, rather than being unphysical as initially thought, represented a class of hitherto unobserved and unimagined particles—antimatter. The existence of particles of antimatter was confirmed with the discovery of the positron² (or anti-electron) by Anderson in 1932, but it is still unknown why matter, rather than antimatter, survived after the Big Bang. As a result, experimental studies of antimatter^{3–7}, including tests of fundamental symmetries such as charge–parity and charge–parity–time, and searches for evidence of primordial antimatter, such as antihelium nuclei, have high priority in contemporary physics research. The fundamental role of the hydrogen atom in the evolution of the Universe and in the historical development of our understanding of quantum physics makes its antimatter counterpart—the antihydrogen atom—of particular interest. Current standard-model physics requires that hydrogen and antihydrogen have the same energy levels and spectral lines. The laser-driven 1S–2S transition was recently observed⁸ in antihydrogen. Here we characterize one of the hyperfine components of this transition using magnetically trapped atoms of antihydrogen and compare it to model calculations for hydrogen in our apparatus. We find that the shape of the spectral line agrees very well with that expected for hydrogen and that the resonance frequency agrees with that in hydrogen to about 5 kilohertz out of 2.5×10^{15} hertz. This is consistent with charge–parity–time invariance at a relative precision of 2×10^{-12} —two orders of magnitude more precise than the previous determination⁸—corresponding to an absolute energy sensitivity of 2×10^{-20} GeV.

The transition of interest here, between the ground state and the first excited state of antihydrogen, has an energy of about 10.2 eV. The frequency of this transition in hydrogen has been measured⁸ to a few parts in 10^{15} . We previously demonstrated⁷ the existence of the transition in antihydrogen, localizing the frequency to a few parts in 10^{10} . Here we characterize the spectral line shape of the transition to the limits of precision of our current apparatus.

Matter and antimatter annihilate each other, so antihydrogen must be synthesized and then held in ultrahigh vacuum, in isolation from matter, to be studied. The ALPHA-2 apparatus at CERN (Fig. 1) combines antiprotons from the antiproton decelerator⁹ with positrons from a positron accumulator^{10,11} to produce and trap¹² atoms of antihydrogen. Antihydrogen can be trapped in ALPHA-2's magnetic multipole trap if

it is produced with a kinetic energy of less than 0.54 K in temperature units. The techniques that we use to produce antihydrogen that is cold enough to trap are described elsewhere^{12–14}. In round numbers, a typical trapping trial in ALPHA-2 involves mixing 90,000 antiprotons with 3,000,000 positrons to produce 50,000 antihydrogen atoms, about 20 of which will be trapped. The anti-atoms are confined by the interaction of their magnetic moments with the inhomogeneous magnetic field. The cylindrical trapping volume for antihydrogen has a diameter of 44.35 mm and a length of 280 mm.

The key to anti-atomic spectroscopy, as developed so far^{7,15,16}, is to illuminate a sample of trapped antihydrogen atoms with electromagnetic radiation (microwaves or laser photons) that causes atoms to be lost from the trap if the radiation is on resonance with the transition of interest. ALPHA-2's silicon vertex detector¹⁷ (Fig. 1) affords us single-atom detection capability for the annihilation events associated with lost antihydrogen atoms or antiprotons that encounter the walls of the apparatus. The silicon vertex detector tracks the charged pions from the antiproton annihilation, and various reconstruction algorithms are used to determine the location (vertex) of each annihilation and to distinguish antiprotons from cosmic-ray background using multivariate analysis¹⁸ (Methods).

To excite the 1S–2S transition, we use a cryogenic, in vacuo enhancement cavity (Fig. 1) for continuous-wave light from a 243-nm laser system (Methods) to boost the intensity in the trapping volume. Long interaction times are possible, because the anti-atoms have a storage lifetime of at least 60 h in the trap. Two counter-propagating photons can resonantly excite the ground-state atoms to the 2S state. Absorption of a third photon ionizes the atom, leading to loss of the antiproton from the trap. Atoms that decay from the 2S to the 1S state via coupling to the 2P state may also be lost, owing to a positron spin-flip¹⁹.

Referring to the energy-level diagram of hydrogen in Fig. 2, there are two trappable, hyperfine substates of the 1S ground state (labelled 'c' and 'd'). In practice, we find that these states are, on average, equally populated in our trap: $N_c = N_d = N_i/2$, where N_i is the number of ground-state atoms that are initially trapped in an experimental trial. The 2S state has corresponding hyperfine levels, and we refer to the transitions between the two manifolds as d–d (Fig. 2) and c–c (not pictured).

For each experimental trial, we first accumulate antihydrogen atoms from three mixing cycles or 'stacks'¹³ and then remove any leftover charged particles using pulsed electric fields. After a wait of about 10 s

¹Department of Physics, University of Liverpool, Liverpool, UK. ²Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark. ³Department of Physics, College of Science, Swansea University, Swansea, UK. ⁴School of Physics and Astronomy, University of Manchester, Manchester, UK. ⁵Cockcroft Institute, Sci-Tech Daresbury, Warrington, UK. ⁶TRIUMF, 4004 Wesbrook Mall, Vancouver, British Columbia, Canada. ⁷Department of Physics, University of California at Berkeley, Berkeley, CA, USA. ⁸Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. ⁹Department of Physics, Ben-Gurion University of the Negev, Beer-Sheva, Israel. ¹⁰Department of Physics and Astronomy, University of Calgary, Calgary, Alberta, Canada. ¹¹Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada. ¹²Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada. ¹³Department of Physics, Stockholm University, Stockholm, Sweden. ¹⁴Department of Physics and Astronomy, York University, Toronto, Ontario, Canada. ¹⁵Department of Physics and Astronomy, University of Victoria, Victoria, British Columbia, Canada. ¹⁶Department of Physics and Astronomy, Purdue University, West Lafayette, IN, USA. ¹⁷Soreq NRC, Yavne, Israel. ¹⁸Physics Department, Marquette University, Milwaukee, WI, USA. ¹⁹IRFU, CEA/Saclay, Gif-sur-Yvette Cedex, France. *e-mail: jeffrey.hangst@cern.ch

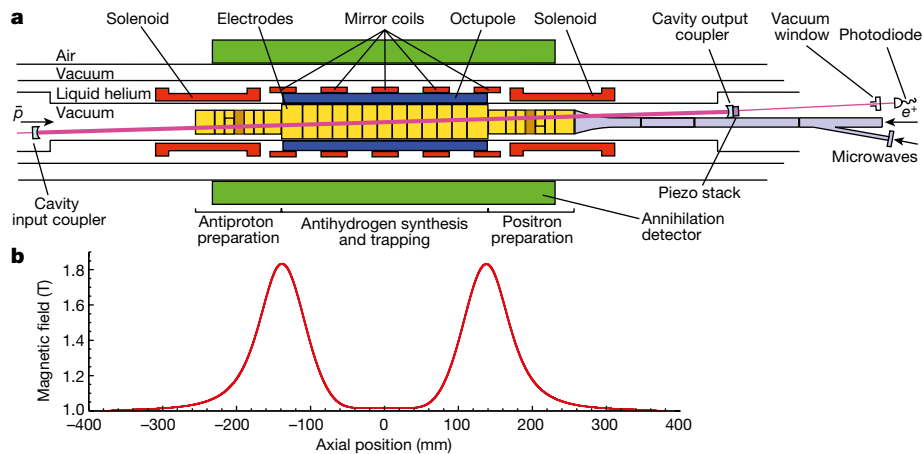


Fig. 1 | The ALPHA-2 central apparatus and magnetic field profile. **a, b**, Penning traps, comprising stacks of cylindrical electrodes immersed in a uniform axial magnetic field generated by an external solenoid (not shown), are used to confine and manipulate antiprotons (\bar{p}) and positrons (e^+) to produce antihydrogen. Cold (less than 0.5 K) anti-atoms can be trapped radially by the octupole field and axially by the magnetic well that is formed by the five mirror coils and plotted in **b**. The 243-nm laser light is injected from the antiproton side (left in **a**) and is aligned and position-stabilized on the fixed optical cavity axis. The laser beam crosses the trap

axis at an angle of 2.3° . The piezoelectric actuator behind the output coupler is used to modulate the cavity length to lock the cavity to the laser frequency. The axial scale in **a** and **b** is the same; the radial extent of the annihilation detector is larger than illustrated. The vacuum window and photo-diode are further to the right (by about 1 m) than illustrated. The brown-shaded electrodes are used to apply blocking potentials during the experimental trials to ensure that antiprotons that result from ionization are confined to annihilate in the active volume of the detector⁷.

to allow any excited atoms to decay to the ground state, the trapped population is exposed to laser radiation at a fixed frequency for 300 s. The frequencies used here were chosen to probe only the d–d transition (Fig. 2). Following the laser exposure, we use microwave radiation to remove the $1S_c$ state atoms by driving a resonant spin-flip^{15,16}. The microwave frequency is scanned over 9 MHz in 32 s; these parameters and the injected power level (160 mW at the vacuum feed-through) are chosen to eject anti-atoms quickly while minimizing the perturbation of the vacuum and cryogenic environment. The silicon vertex detector is used to detect annihilations of antihydrogen atoms that are lost during the laser and microwave exposures. Finally, the atom-trap magnets are ramped down in 1.5 s, so that any surviving anti-atoms would be released and their annihilations detected. If the microwave removal of $1S_c$ -state atoms is 100% effective, then the surviving particles would be only $1S_d$ -state atoms that were not removed by laser action.

Table 1 | Antihydrogen atom counts

	Laser detuning, D (kHz)	Number of trials	Atoms lost during laser exposure, L	Atoms lost during microwave exposure, M	Surviving atoms, S	Initially trapped atoms, N_i
Set 1	–200	21	7 ± 7	383 ± 23	504 ± 25	894 ± 35
	–100	21	22 ± 9	415 ± 24	494 ± 24	931 ± 35
	0	21	264 ± 24	423 ± 24	217 ± 16	904 ± 38
	+100	21	75 ± 14	411 ± 23	424 ± 23	910 ± 35
Set 2	–200	21	26 ± 9	394 ± 23	466 ± 24	886 ± 34
	–25	21	113 ± 16	423 ± 24	326 ± 20	862 ± 35
	0	21	219 ± 22	390 ± 23	269 ± 18	878 ± 37
	+25	21	173 ± 20	438 ± 24	296 ± 19	907 ± 37
Set 3	–200	23	8 ± 7	354 ± 22	479 ± 24	841 ± 33
	0	23	303 ± 26	454 ± 25	248 ± 17	$1,005 \pm 40$
	+50	23	176 ± 20	390 ± 23	339 ± 20	905 ± 37
	+200	23	36 ± 11	446 ± 24	459 ± 23	941 ± 35
Set 4	–200	21	7 ± 7	525 ± 26	541 ± 25	$1,073 \pm 37$
	–50	21	86 ± 15	475 ± 25	495 ± 24	$1,056 \pm 38$
	0	21	274 ± 25	480 ± 25	275 ± 18	$1,029 \pm 40$
	+25	21	202 ± 21	516 ± 26	305 ± 19	$1,023 \pm 38$
Total		344	1,991	6,917	6,137	15,045

The integrated number of antihydrogen atoms is listed for each laser detuning (at 243 nm) within each set of trials. The background has been subtracted. Uncertainties quoted are one standard deviation (s.d.) counting errors. We refer to L as the ‘appearance signal’; S is used to infer the ‘disappearance signal’.

We collected data for nine different laser frequencies in four sets. Each set involved four distinct frequencies and 21 (or 23, see below) trials at each of these frequencies. In each set, two of the frequencies were always the calculated hydrogen on-resonance frequency at zero laser power (zero detuning) and a far-off-resonance frequency (–200 kHz detuning at 243 nm), as used previously⁷. The other two frequencies in each set were chosen to address various detunings in the neighbourhood of the d–d resonance. The data are summarized in Table 1. The repetition of the points at –200 kHz and zero detuning was intended to address variations in laser power and trapping number between sets. The repetition at +25 kHz was a check of reproducibility. During the accumulation of data for each set, the four frequencies were interleaved in a varying order and the operators were blinded as to the identity of each frequency setting. The power of the enhancement cavity (about 1 W) was monitored by measuring the transmitted power outside of the vacuum chamber (Fig. 1). Each set was preceded by a thermal cycle of the apparatus to regenerate the cryo-pumping surface.

The background-corrected numbers in Table 1 are calculated from raw detector events using the measured, overall efficiencies of the silicon vertex detector. These efficiencies depend on the particular multivariate analysis algorithm that was used to distinguish antiproton annihilations from cosmic rays (Methods) in the relevant time window. The efficiencies and background rates are listed in Table 2.

The number of initially trapped atoms N_i for a trial is unknown a priori, but was typically about 60 at the beginning of a measurement set. In Table 1, the total number of atoms for each group of trials is assumed to be the sum $L + M + S$ of the numbers of atoms lost during laser (L) or microwave (M) exposure and the number of surviving atoms (S) (see Table 1). The trapping rate declined slowly but reproducibly during each set (Extended Data Fig. 1). The third set has 23 trials at each

Table 2 | Annihilation detector efficiencies and background rates

	Efficiency	Uncertainty	Background rate (10^{-3} s^{-1})	Uncertainty (10^{-3} s^{-1})
Laser exposure (300 s)	0.472	0.001	1.04	0.11
Microwave exposure (32 s)	0.801	0.002	33.0	0.6
Release of surviving atoms (1.6 s)	0.852	0.002	191	1

The detection efficiencies and background rates of the silicon vertex detector, as determined by the multivariate analysis (Methods), are listed for the three observation windows. The 1.6-s window during which the surviving atoms are released extends for 0.1 s after the magnet rampdown is complete.

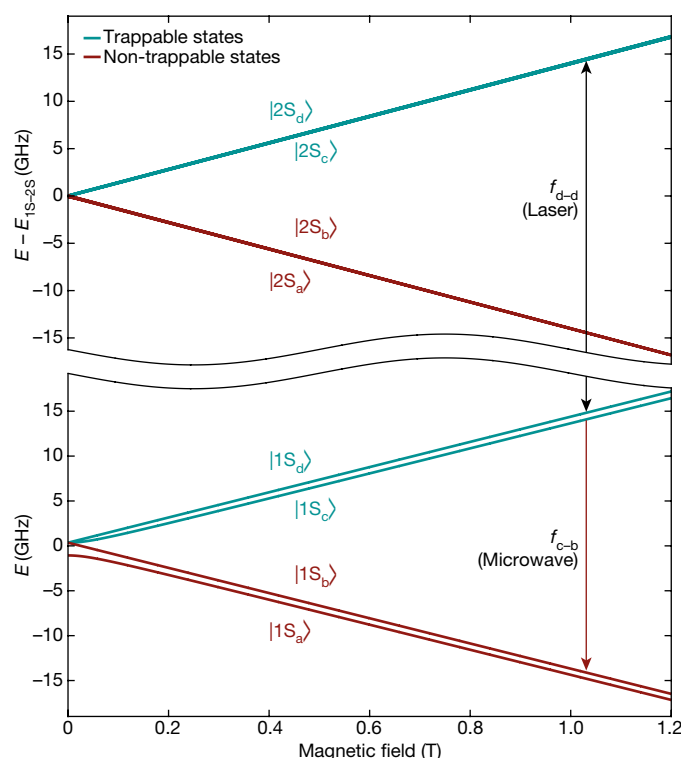


Fig. 2 | Hydrogenic energy levels. Calculated energies (E ; for hydrogen) of the hyperfine sublevels of the 1 S (bottom) and 2 S (top) states are plotted against magnetic field strength. The centroid energy difference $E_{1S-2S} = 2.4661 \times 10^{15}$ Hz has been suppressed on the vertical axis. The vertical black arrow indicates the two-photon laser transition probed here (frequency f_{d-d}); the red arrow illustrates the microwave transition used to remove the 1S_c state atoms (frequency f_{c-b}).

frequency because of a hardware failure in an early block of four trials; extra trials were added to compensate for the excluded data.

To examine the general features of the measurement results, we plot (Fig. 3a) the four datasets on one graph by using a simple scaling. The points at zero (on-resonance) and -200 -kHz detuning (at which no signal is expected⁷), repeated for each set, are used for the scaling. For the laser exposure ('appearance') data, we define a scaled response at detuning D within each set: $r_l(D) = L(D)/L(0)$. Similarly, for the surviving population ('disappearance' data), we use $r_s(D) = [S(-200 \text{ kHz}) - S(D)]/[S(-200 \text{ kHz}) - S(0)]$. The uncertainties shown are due to Poissonian counting errors only. For comparison, we also plot the results of a simulation¹⁹ based on the expected behaviour of hydrogen in our trap for a cavity power of 1 W, scaled to the zero-detuning data point. We see that the peak position and the width of the scaled spectral line are consistent with the calculation for hydrogen and that the experiment generally reproduces the predicted asymmetric line shape. There is also good agreement between the appearance and disappearance data (Fig. 3a).

The simulation involves propagating the trapped atoms in an accurate model of the magnetic trap. When an atom crosses the laser beam, which has a waist of $200 \mu\text{m}$ at the cavity centre, we calculate the two-photon excitation probability, taking into account transit-time broadening, the a.c. Stark shift and the residual Zeeman effect. The simulation determines whether excited atoms are lost owing to ionization or to a spin-flip event. The variable input parameters for the simulation are the cavity power and the laser frequency. The modelled response is asymmetric in frequency owing to the residual Zeeman effect¹⁹. The width of the line, for our experimental parameters, is dominated by transit-time broadening, which contributes about 50 kHz full-width at half-maximum (FWHM) at 243 nm . For 1 W of cavity power, the a.c. Stark shift is about 2.5 kHz to higher frequency and the ionization contributes about 2 kHz to the natural line width.

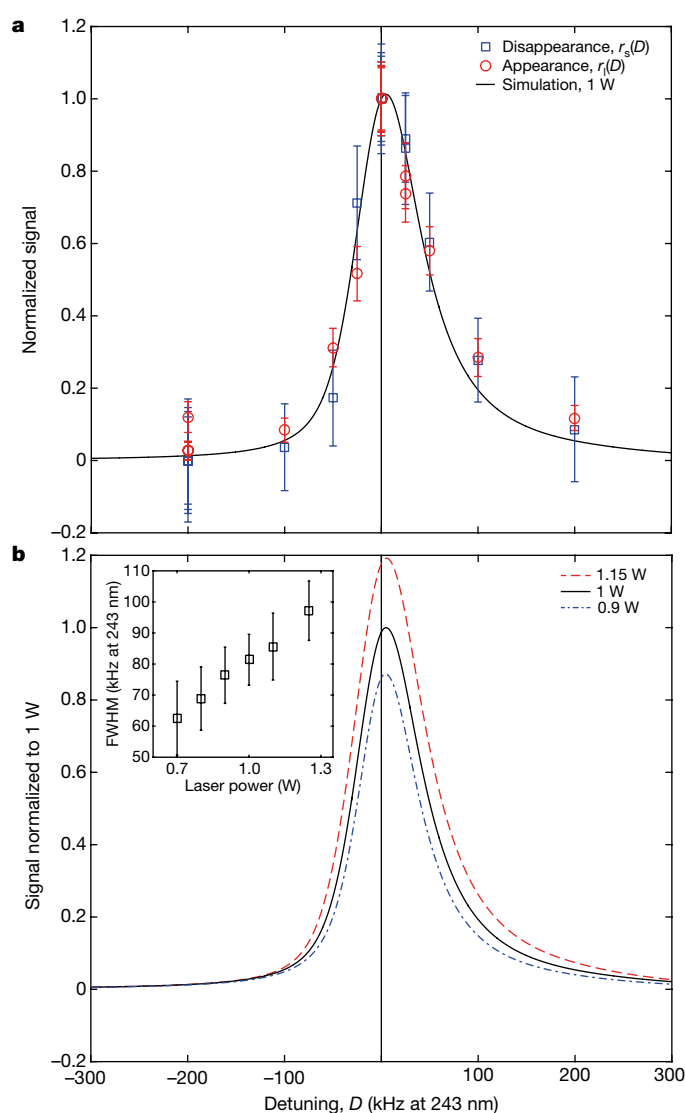


Fig. 3 | Spectral line of antihydrogen. **a**, The complete dataset, scaled as described in the text. The simulated curve (not a fit, drawn for qualitative comparison only) is for a stored cavity power of 1 W and is scaled to the data at zero detuning. 'Appearance' refers to annihilations that are detected during laser irradiation; 'disappearance' refers to atoms that are apparently missing from the surviving sample. The error bars are 1-s.d. counting uncertainties. **b**, Three simulated line shapes (for hydrogen) are depicted for different cavity powers to illustrate the effect of power on the size and the frequency at the peak. The width of the simulated line (FWHM) as a function of laser power is plotted in the inset.

To make a more quantitative comparison of the experimental results with the expectations for hydrogen, it is necessary to scrutinize differences between the four datasets. The overall response should be linear in the number of atoms addressed, so it is possible to normalize for this. However, the line width depends on the stored power in the cavity, as does the frequency of the peak (Fig. 3b). The cavity power is difficult to measure in our geometry because the amount of transmitted light depends sensitively on the small transmission from the output coupler (about 0.05%) and on absorption in the optical elements through which the transmitted light exits (Fig. 1). We observe that the transmitted power can degrade, owing to accumulated ultraviolet damage to the window and mirror substrate, whereas the finesse of the cavity does not change.

A modelling approach that self-consistently accounts for fluctuations in experimental parameters is a simultaneous fit in which we allow the four sets to have distinct powers (P_{1-4}), but the same frequency shift with respect to the hydrogen calculation (Methods). We require that

Table 3 | Summary of uncertainties

Type of uncertainty	Estimated size (kHz)	Comment
Statistical uncertainties	3.8	Poisson errors and curve fitting to measured data
Modelling uncertainties	3	Fitting of simulated data to piecewise-analytic function
Modelling uncertainties	1	Waist size of the laser, antihydrogen dynamics
Magnetic-field stability	0.03	From microwave removal of 1S _c -state atoms (see text)
Absolute magnetic-field measurement	0.6	From electron cyclotron resonance
Laser-frequency stability	2	Limited by GPS clock
d.c. Stark shift	0.15	Not included in simulation
Second-order Doppler shift	0.08	Not included in simulation
Discrete frequency choice of measured points	0.36	Determined from fitting sets of pseudo-data
Total	5.4	

The estimated statistical and systematic errors (at 121 nm) are tabulated.

the average powers for the appearance and disappearance data within a set are the same. We find the parameters that best reproduce the data to be: $P_1 = 1135(50)$ mW, $P_2 = 904(30)$ mW, $P_3 = 1123(43)$ mW, $P_4 = 957(31)$ mW and $\delta f = -0.44 \pm 1.9$ kHz, where δf is the difference (at 243 nm) between the resonant frequency inferred from the fit and the resonant frequency of hydrogen expected for our system, both at zero power. The uncertainties represent the 68% confidence interval of a least-squares fit and do not take into account systematic uncertainties. The fit uses the five variables identified above, and the individual data points at each frequency are weighted by their Poissonian counting errors. We include an uncertainty of 3.8 kHz (Table 3) in the final resonance frequency to represent statistical and curve-fitting uncertainties.

Considering systematic effects, the microwave removal procedure for the 1S_c-state atoms provides a reproducibility check on the strength of the magnetic field at the centre of the trap. At the beginning of each data-taking shift, the magnetic field of the external solenoid magnet was reset to a standard value using an electron cyclotron resonance technique¹⁶. For the complete dataset, we find that the variations in the magnetic field at the minimum field of about 1 T are about 3.2×10^{-5} T (1 s.d.). This corresponds to a resonance frequency shift¹⁹ of only about 15 Hz at 243 nm for the d–d transition. (At 1 T, the c–c transition is about 20 times more sensitive to magnetic field shifts, which is why the d–d transition is more attractive here.) The laser frequency was tuned with respect to the minimum of the magnetic well, such that the resonance condition should be met in the centre of the trap for zero detuning in the limit of zero laser power. The accuracy of the magnetic-field determination corresponds to an uncertainty of 300 Hz in the 243-nm laser frequency.

Including all of the statistical and systematic uncertainties that we have identified (Table 3, for 121 nm), our fit of the experimental data to the hydrogen model yields

$$f_{d-d} = 2,466,061,103,079.4(5.4) \text{ kHz}$$

The value (Methods) for hydrogen calculated at the minimum field in our system (1.03285(63) T) is

$$f_{d-d} = 2,466,061,103,080.3(0.6) \text{ kHz}$$

where the uncertainty is determined by the experimental error in measuring the field.

Owing to the motion of the antihydrogen atoms in the inhomogeneous trapping field, this comparison is necessarily model-dependent. We therefore conclude that the measured resonance frequency for this transition in antihydrogen is consistent with the expected hydrogen frequency to a precision of about 2×10^{-12} . Although the precision of our measurement is still a few orders of magnitude short of the state of the art with a cold hydrogen beam⁸, the modern frequency reference permits the accuracy of our experiment to exceed that achieved with

trapped hydrogen²⁰ as recently as the mid-1990s. We used a total of about 15,000 antihydrogen atoms to obtain this result, compared to 10^{12} trapped atoms in the analogous matter experiment. Our dataset was accumulated over a period of ten weeks, illustrating that the antihydrogen trapping procedure is robust and that systematic effects are manageable. ALPHA's emergent antihydrogen production, storage and detection techniques, together with advances in ultraviolet laser technology and frequency metrology, pioneered by Hänsch and colleagues, enable precision anti-atom spectroscopy.

Precision experiments at the antiproton decelerator have recently constrained the properties of the antiproton through studies in Penning traps^{21,22} or with antiprotonic helium²³. For example, the antiproton charge-to-mass ratio is known to agree with that of the proton to 69 parts per trillion²¹, equivalent to an energy sensitivity of 9×10^{-27} GeV. The ratio of the antiproton mass to the electron mass has been shown to agree with its proton counterpart²³ to 8×10^{-10} , and antihydrogen has been shown to be neutral²⁴ to 0.7 parts per billion. Our measurement of antihydrogen probes different and complementary physics at a precision of a few parts per trillion, or an energy level of 2×10^{-20} GeV. This already exceeds the precision (4×10^{-19} GeV) in the mass difference of neutral kaons and antikaons²⁵, which has long been the standard for particle-physics tests of charge–parity–time invariance.

Near-term improvements in the ALPHA-2 apparatus will include a larger waist size for the radiation in the optical cavity to reduce transit-time broadening, operation at lower magnetic fields and operational improvements to accelerate data acquisition and to reduce statistical uncertainties. Future measurements will require an upgrade to our frequency reference to exceed a fractional precision of 8×10^{-13} (Methods). The rapid progress detailed here confirms that, in principle, there is nothing to prevent the achievement of hydrogen-like precision in antihydrogen and the associated very sensitive test of charge–parity–time symmetry in this system.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0017-2>.

Received: 11 December 2017; Accepted: 6 March 2018;

Published online: 04 April 2018

1. Dirac, P. A. M. The quantum theory of the electron. *Proc. R. Soc. Lond. A* **117**, 610–624 (1928).
2. Anderson, C. D. The positive electron. *Phys. Rev.* **43**, 491–494 (1933).
3. Bertsche, W. A., Butler, E., Charlton, M. & Madsen, N. Physics with antihydrogen. *J. Phys. B* **48**, 232001 (2015).
4. Hori, M. & Walz, J. Physics at CERN's Antiproton Decelerator. *Prog. Part. Nucl. Phys.* **72**, 206–253 (2013).
5. The LHCb Collaboration. Measurement of matter–antimatter differences in beauty baryon decays. *Nature Physics* **13**, 391–396 (2017).
6. The AMS Collaboration. Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the Alpha Magnetic Spectrometer on the International Space Station. *Phys. Rev. Lett.* **117**, 091103 (2016).
7. Ahmadi, M. et al. Observation of the 1S–2S transition in antihydrogen. *Nature* **541**, 506–510 (2017).
8. Parthey, C. G. et al. Improved measurement of the hydrogen 1S–2S transition frequency. *Phys. Rev. Lett.* **107**, 203001 (2011).
9. Maury, S. The antiproton decelerator: AD. *Hyperfine Interact.* **109**, 43–52 (1997).
10. Murphy, T. J. & Surko, C. M. Positron trapping in an electrostatic well by inelastic collisions with nitrogen molecules. *Phys. Rev. A* **46**, 5696–5705 (1992).
11. Surko, C. M., Greaves, R. G. & Charlton, M. Stored positrons for antihydrogen production. *Hyperfine Interact.* **109**, 181–188 (1997).
12. Andresen, G. B. et al. Trapped antihydrogen. *Nature* **468**, 673–676 (2010).
13. Ahmadi, M. et al. Antihydrogen accumulation for fundamental symmetry tests. *Nat. Commun.* **8**, 681 (2017).
14. Andresen, G. B. et al. Evaporative cooling of antiprotons to cryogenic temperatures. *Phys. Rev. Lett.* **105**, 013003 (2010).
15. Ahmadi, M. et al. Observation of the hyperfine spectrum of antihydrogen. *Nature* **548**, 66–69 (2017).
16. Amole, C. et al. Resonant quantum transitions in trapped antihydrogen atoms. *Nature* **483**, 439–443 (2012).

17. Amole, C. et al. Silicon vertex detector upgrade in the ALPHA experiment. *Nucl. Instrum. Methods A* **732**, 134–136 (2013).
18. Hoecker, A. et al. TMVA - toolkit for multivariate data analysis. Preprint at <https://arxiv.org/abs/physics/0703039> (2007).
19. Rasmussen, C. Ø., Madsen, N. & Robicheaux, F. Aspects of 1S–2S spectroscopy of trapped antihydrogen atoms. *J. Phys. B* **50**, 184002 (2017); corrigendum, <https://doi.org/10.1088/1361-6455/aab1a9> (2018).
20. Cesar, C. L. et al. Two-photon spectroscopy of trapped atomic hydrogen. *Phys. Rev. Lett.* **77**, 255–258 (1996).
21. Ulmer, S. et al. High-precision comparison of the antiproton-to-proton charge-to-mass ratio. *Nature* **524**, 196–199 (2015).
22. Smorra, C. et al. A parts-per-billion measurement of the antiproton magnetic moment. *Nature* **550**, 371–374 (2017).
23. Hori, M. et al. Buffer-gas cooling of antiprotonic helium to 1.5 to 1.7 K, and antiproton-to-electron mass ratio. *Science* **354**, 610–614 (2016).
24. Ahmadi, M. et al. An improved limit on the charge of antihydrogen from stochastic acceleration. *Nature* **529**, 373–376 (2016).
25. Patrignani, C. et al. Review of particle physics. *Chin. Phys. C* **40**, 100001 (2016).

Acknowledgements All authors are members of the ALPHA Collaboration. This work was supported by: the European Research Council through its Advanced Grant programme (J.S.H.); CNPq, FAPERJ, RENAFÉ (Brazil); NSERC, NRC/TRIUMF, EHPDS/EHDRS, FQRNT (Canada); FNU (Nice Centre), Carlsberg Foundation (Denmark); ISF (Israel); STFC, EPSRC, the Royal Society and the Leverhulme Trust (UK); DOE, NSF (USA); and VR (Sweden). We are grateful for the efforts of the CERN Antiproton Decelerator team, without which these experiments could not have taken place. We thank J. Tonoli (CERN) and his staff for extensive, time-critical help with machining work. We thank the staff of the Superconducting Magnet Division at Brookhaven National Laboratory for collaboration and fabrication of the trapping magnets. We thank C. Marshall (TRIUMF) for his work on the ALPHA-2 cryostat. We thank D. Tommasini and A. Milanese (CERN) for the fabrication of conventional magnets for ALPHA-2. We acknowledge the influence of T. Hänsch on the methodology and the hardware used here. J.S.H. and C.L.C. acknowledge early encouragement from D. Kleppner (MIT) in these endeavours. We thank F. Besenbacher (Aarhus) for support in procuring the ALPHA-2 external solenoid. We thank J. Serrano (CERN) for making the atomic clock available to us.

Reviewer information *Nature* thanks D. Horvath, K. Jungmann and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions This experiment was based on data collected using the ALPHA-2 antihydrogen trapping apparatus, designed and constructed by the ALPHA Collaboration using methods developed by the entire collaboration. The entire collaboration participated in the operation of the apparatus and the data-taking activities. The laser and internal cavity system was conceived, implemented, commissioned and operated by W.B., N.M., J.S.H., S.E., C.Ø.R., S.A.J., C.L.C., B.X.R.A. and G.S. F.R., C.Ø.R., J.F. and N.M. developed the simulation program for laser interaction with magnetically trapped atoms. Analysis of the spectral line shapes was done by C.Ø.R., N.M. and J.S.H. Detailed analysis of the antiproton annihilation detector data was done by J.T.K.M. and A.O. Implementation of the microwave system and analysis of the microwave data was done by T.F. and M.E.H. The positron accumulator is the responsibility of C.J.B., M.C., C.A.I. and D.P.v.d.W. The manuscript was written by J.S.H., N.M., C.Ø.R., S.A.J. and J.T.K.M., with help from A.O., C.L.C. and S.E. The manuscript was then edited and improved by the entire collaboration.

Competing interests The authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41586-018-0017-2>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to J.S.H.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

Time evolution of the dataset. The time evolution of the atoms detected in one of the datasets is depicted in Extended Data Fig. 1.

Laser system for 243-nm light. A Toptica TA-FHG pro laser system uses a pair of frequency-doubling cavities to generate 150 mW of 243-nm light from a 972-nm extended cavity diode laser (ECDL). The 243-nm beam is mode-matched to the 1S–2S enhancement cavity and sent along a 7-m-long path with active beam-pointing stabilization between the laser laboratory and the ALPHA-2 apparatus. The enhancement cavity is locked to the laser frequency using a single piezoelectric actuator located behind the output coupler mirror²⁶ to feedback on an error signal generated via the Pound–Drever–Hall technique²⁷. The light transmitted through the cavity is monitored using a photodiode that is located outside the vacuum system. The cavity has a measured finesse of 250 and achieves a circulating power of approximately 1 W.

The 972-nm ECDL is frequency-stabilized (also using the Pound–Drever–Hall technique) to a Menlo Systems ultralow-expansion cavity via an acousto-optic modulator, which shifts the light from the 1S–2S transition frequency of the laser to the closest resonance of the ultralow-expansion cavity. The resonance frequency of the cavity is monitored continuously using a Menlo Systems femtosecond frequency comb, which is referenced to atomic time using a K + K Messtechnik GPS-disciplined quartz oscillator.

The measured difference between the ultralow-expansion resonance frequency and a comb line with a known frequency is fed forward to the control of the acousto-optic modulator with an averaging time of 20 s to remove long-term drifts. The uncertainty of the frequency difference over the 20-s averaging period corresponds to an Allan deviation²⁸ of 75 Hz at 972 nm (300 Hz at 243 nm). One of the frequency-comb counters is used to measure the signal from a Symmetricom CS4000 caesium clock to confirm correct operation of the quartz oscillator and the radio-frequency chain of the frequency comb. The count reaches a fractional Allan deviation of 8×10^{-13} after 1,000 s of averaging, which corresponds to fluctuations of 250 Hz at 972 nm (1 kHz at 243 nm).

An independent, identical, 972-nm ECDL frequency stabilized to an independent, identical, ultralow-expansion cavity is used to evaluate the short-term line width of the spectroscopy laser. The beat note generated between the two 972-nm lasers has a spectrum composed of individual lines, each with a line width of less than 1 Hz, within a 300-Hz (1.2 kHz at 243 nm) FWHM Gaussian envelope. The source of the broadening is thought to be acoustic noise within the laser laboratory; work is ongoing to reduce the broadening effect.

Suppression of cosmic-ray background. To determine the signal events in the (a) 1.6-s, (b) 32-s and (c) 300-s observation windows, we require three different suppression techniques. (The 1.6-s window extends to 0.1 s after the magnet rampdown is complete.) We tune the multivariate analysis (MVA) for each of the three windows to optimize the statistical significance of the estimated signal. Annihilation events are distinguished from background events (primarily cosmic rays) by their distinctive topologies. Fourteen selection variables that are sensitive to the difference between annihilation and background events were used as inputs to an MVA package¹⁸. The variables included are: (i) the total number of channels registering ‘hits’ by charged particles; (ii) the radial coordinates of the reconstructed annihilation vertex; (iii) the sum of the squared residual distances of hits from a fitted straight line; six topological variables (iv–ix); and five additional variables (x–xiv). The topological variables are: (iv) a sphericity variable; (v) the cosine of the angle between the event axis and the detector axis; (vi) the angle between the event axis and the vertical direction in the x – y plane; (vii) the number of reconstructed tracks; (viii) the number of three-hit combinations used as track candidates; (ix) the distance of closest approach of the tracks. The additional variables are: (x) the minimum and (xi) mean values of the track radius in canonical form; (xii) the minimum and (xiii) mean values of the pitch of the helical track in canonical form; and (xiv) an integer sum of the sense of curvature (left = −1 or right = +1) for all of the tracks in the event.

The signal data and background data used for MVA training and testing comprise a set of 580,846 annihilation events and 3,740,613 background events. The signal events were produced during antiproton and positron mixing in the apparatus and contain less than 1% background. Background events were collected during times when there were no antiprotons in the apparatus.

The 1.6-s observation window. A classifier cut was chosen to optimize the significance for an expected 200 counts of signal and 350 counts of background. The analysis gives a background rate of $0.191 \pm 0.001 \text{ s}^{-1}$ and an efficiency of 0.852 ± 0.002 (statistical error only) annihilations per detector trigger.

The 32-s observation window. The analysis was chosen to optimize the significance for an expected 400 counts of signal and 3,500 counts of background. The analysis gives a background rate of $0.033 \pm 0.0006 \text{ s}^{-1}$ and an efficiency of 0.801 ± 0.002 (statistical error only) annihilations per detector trigger.

The 300-s observation window. A classifier cut was chosen to optimize the significance for an expected 250 counts of signal and 330,000 counts of background. The analysis gives a background rate of $0.0010 \pm 0.0001 \text{ s}^{-1}$ and an efficiency of 0.472 ± 0.001 (statistical error only) annihilations per detector trigger.

Fitting the data using the hydrogen simulation. The build-up of laser power in the enhancement cavity is one of the primary experimental parameters that influence the data in Table 1. The main effect of a change in laser power is on the amplitude of the measured line, but there is also an effect on the peak position through the a.c. Stark shift and on the line width owing to depletion effects. In our set-up, there is considerable uncertainty in measuring the absolute intra-cavity laser power; relative measurements show that although the constancy of laser power within any single measurement set is good, there are variations between the sets.

To reflect this experimental reality in our analysis of the data, the χ^2 statistic for the full dataset is minimized with respect to a function that, aside from an overall frequency shift, allows a unique laser power in each set and incorporates the effects of those laser powers on the amplitude, line width and line centre based on the simulation of hydrogen in our experiment.

The construction of the fit function therefore starts by running a detailed simulation of hydrogen in the ALPHA-2 magnetic trap for an array of input laser powers and frequencies that spans the experimentally relevant values, in this case from −200 kHz to +300 kHz in laser detuning and from 0.7 W to 1.25 W in laser power. We simulate a total of 365,000 atoms in this array, after which we interpolate to obtain continuous values in both laser detuning and power. The interpolation in power is a linear regression at each detuning in the array, based on the observed linear behaviour. For interpolation in detuning, a fit to a piecewise-analytic function that provides a good approximation to the simulation data is used. An error associated with this fit is included in Table 3. The discrete simulated points and the smooth interpolation are plotted in Extended Data Fig. 2.

Calculation of the resonant frequency for hydrogen. The frequency f_{d-d} is calculated from corrections to the centroid-to-centroid frequency f_{1S2S} :

$$f_{d-d}(B) = f_{1S2S} - \frac{1}{4}[f_{\text{HF}}(1) - f_{\text{HF}}(2)] + [\mu_e(2) - \mu_e(1)]\frac{B}{h} - [\mu_p(2) - \mu_p(1)]\frac{B}{h} + \left(\frac{m}{\mu}\right)^3 \frac{13e^2 a_0^2}{4mh} B^2$$

where h is Planck’s constant, $f_{\text{HF}}(n)$ is the hyperfine splitting of the state with principle quantum number n , μ_e and μ_p are the magnitudes of the magnetic moments of the electron and proton, respectively, μ is the reduced mass of the electron, m is the electron mass, e is the fundamental charge, a_0 is the Bohr radius for an infinite-mass nucleus and B is the magnetic field.

The first correction describes the difference in the hyperfine splittings of the 1S and 2S states. The second (third) correction describes the difference in the magnetic moment of the electron (proton) in these states. The fourth correction describes the difference in the diamagnetic shift.

The magnetic moment of the bound electron is (equation (84))²⁹

$$\mu_e(n) = \mu_e^{\text{free}} \left[1 - \frac{\alpha^2}{3n^2} + \frac{\alpha^4}{2n^3} \left(\frac{1}{2n} - \frac{2}{3} \right) + \frac{\alpha^3}{4\pi n^2} + \frac{\alpha^2}{2n^2} \frac{m}{M} \right]$$

where α is the fine-structure constant, μ_e^{free} is the free-electron dipole moment and M is the proton mass; the dependence on n is described elsewhere^{30,31}. The magnetic moment of the bound proton is (equation (87))²⁹

$$\mu_p(n) = \mu_p^{\text{free}} \left[1 - \frac{\alpha^2}{3n^2} + \frac{\alpha^2}{6n^2} \frac{m}{M} \left(\frac{3 + 4 \times 1.793}{1 + 1.793} \right) \right]$$

where μ_p^{free} is the free-proton dipole moment.

Using current CODATA values³² for the fundamental constants, the frequency is

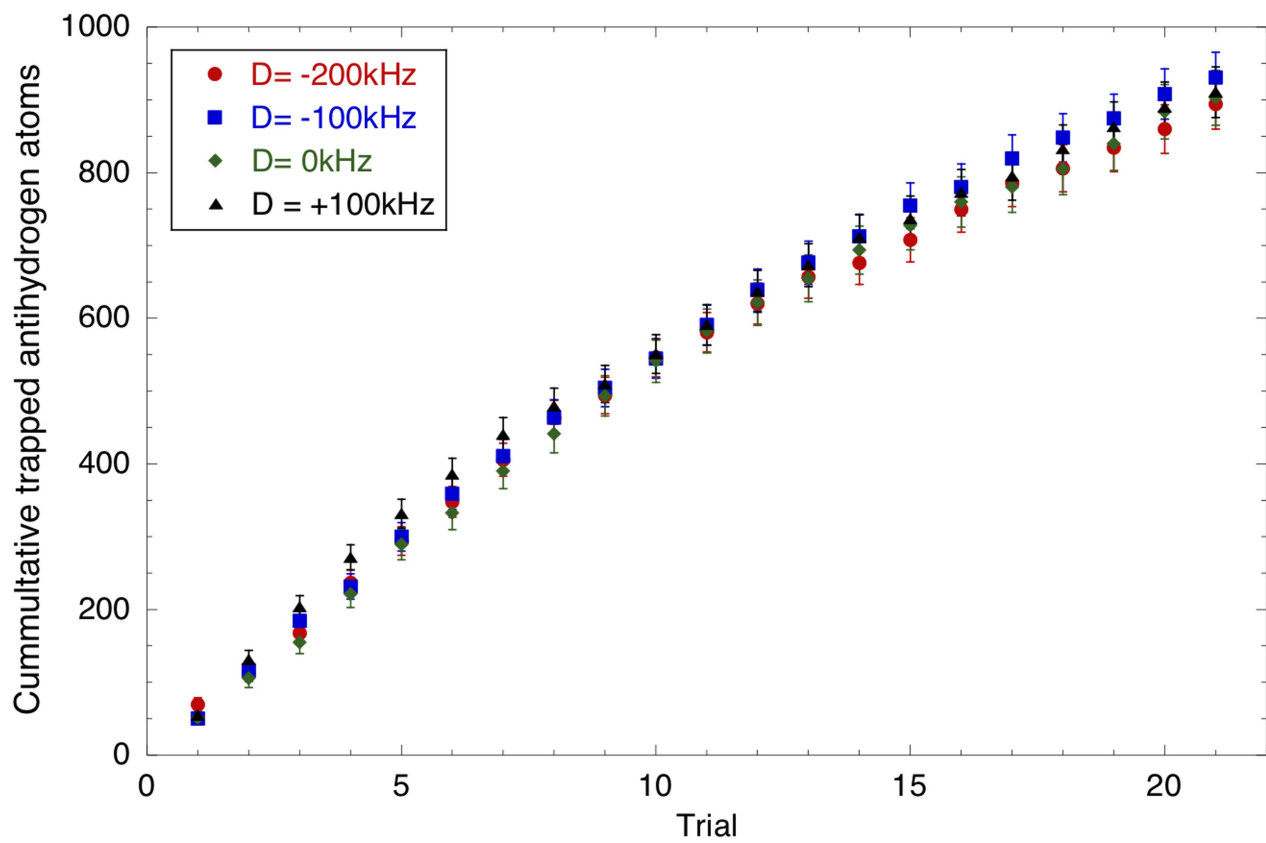
$$f_{d-d} = f_{1S2S} - 310,712.229 \text{ kHz} + 186.071 B \text{ kHz T}^{-1} - 0.283 B \text{ kHz T}^{-1} + 387.678 B^2 \text{ kHz T}^{-2}$$

Sample size. No statistical methods were used to predetermine sample size.

Data availability. The datasets generated and analysed during this study are available from the corresponding author on reasonable request.

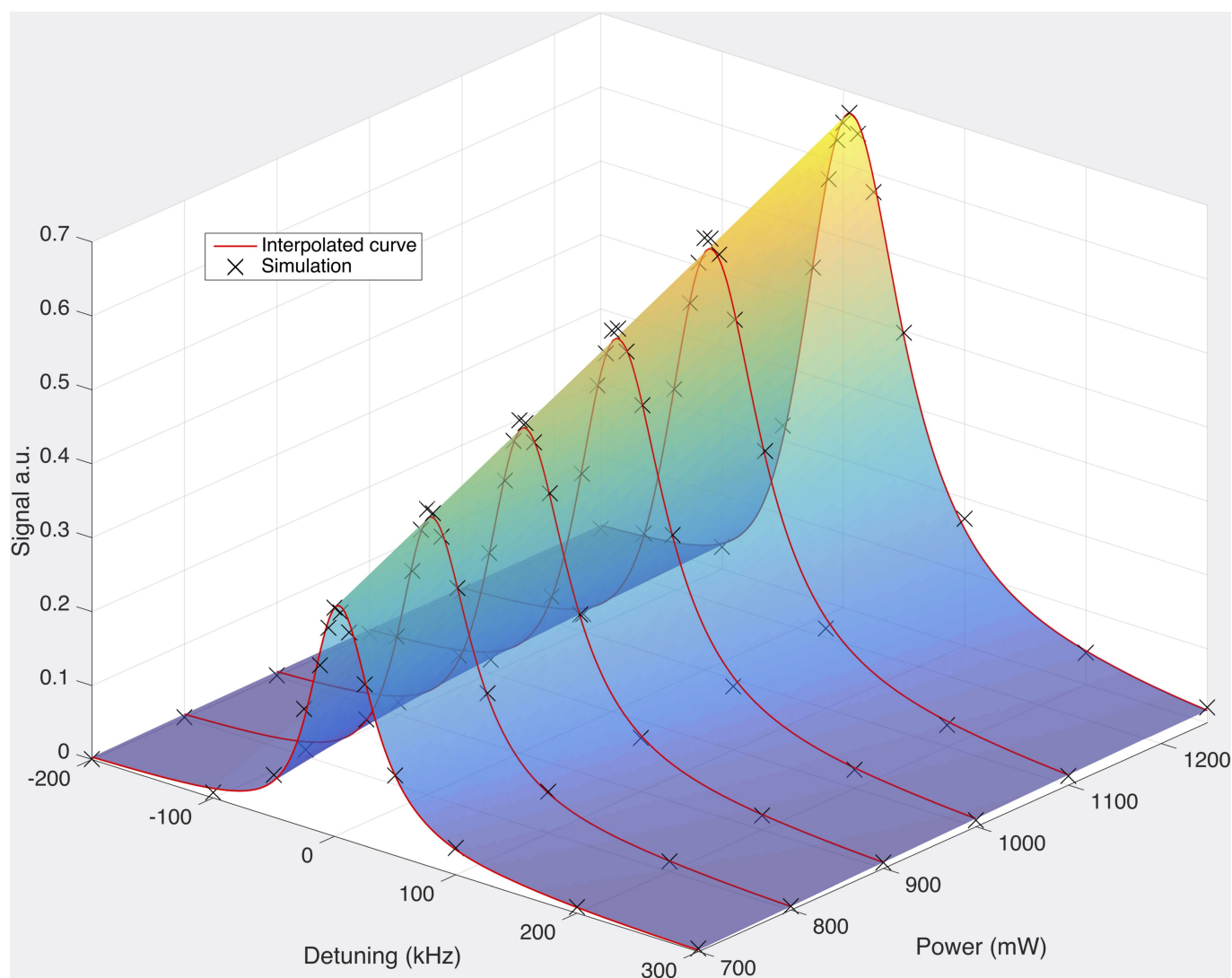
26. Oliveira, A. N. et al. Cryogenic mount for mirror and piezoelectric actuator for an optical cavity. *Rev. Sci. Instrum.* **88**, 063104 (2017).
27. Drever, R. W. P. et al. Laser phase and frequency stabilization using an optical resonator. *Appl. Phys. B* **31**, 97–105 (1983).
28. Allan, D. W. Statistics of atomic frequency standards. *Proc. IEEE* **54**, 221–230 (1966).

29. Mohr, P. J. & Taylor, B. N. CODATA recommended values of the fundamental physical constants: 1998. *Rev. Mod. Phys.* **72**, 351–495 (2000).
30. Shabaev, V. M. QED theory of the nuclear recoil effect on the atomic g factor. *Phys. Rev. A* **64**, 052104 (2001).
31. Jentschura, U. D. et al. Mass measurements and the bound-electron g factor. *Int. J. Mass Spectrom.* **251**, 102–108 (2006).
32. Mohr, P. J., Newell, D. B. & Taylor, B. N. CODATA recommended values of the fundamental physical constants: 2014. *Rev. Mod. Phys.* **88**, 035009 (2016).



Extended Data Fig. 1 | Time evolution of the dataset. The integrated number of atoms is plotted against the trial number for the four detunings

D (-200kHz , -100kHz , 0kHz and 100kHz) in set 1. The error bars are 1-s.d. counting uncertainties.



Extended Data Fig. 2 | Simulation fitting functions. The points (crosses) from the numerical simulation are plotted for various cavity powers. The solid lines represent fits to the simulation by a piecewise-analytic

function. The coloured surface represents the interpolation used to fit the experimental data.

Lightwave valleytronics in a monolayer of tungsten diselenide

F. Langer¹, C. P. Schmid¹, S. Schlauderer¹, M. Gmitra¹, J. Fabian¹, P. Nagler¹, C. Schüller¹, T. Korn¹, P. G. Hawkins², J. T. Steiner², U. Huttner², S. W. Koch², M. Kira^{3*} & R. Huber¹

As conventional electronics approaches its limits¹, nanoscience has urgently sought methods of fast control of electrons at the fundamental quantum level². Lightwave electronics³—the foundation of attosecond science⁴—uses the oscillating carrier wave of intense light pulses to control the translational motion of the electron's charge faster than a single cycle of light^{5–15}. Despite being particularly promising information carriers, the internal quantum attributes of spin¹⁶ and valley pseudospin^{17–21} have not been switchable on the subcycle scale. Here we demonstrate lightwave-driven changes of the valley pseudospin and introduce distinct signatures in the optical readout. Photogenerated electron–hole pairs in a monolayer of tungsten diselenide are accelerated and collided by a strong lightwave. The emergence of high-odd-order sidebands and anomalous changes in their polarization direction directly attest to the ultrafast pseudospin dynamics. Quantitative computations combining density functional theory with a non-perturbative quantum many-body approach assign the polarization of the sidebands to a lightwave-induced change of the valley pseudospin and confirm that the process is coherent and adiabatic. Our work opens the door to systematic valleytronic logic at optical clock rates.

In a crystalline solid, a set of atoms forming the unit cell repeats itself on a periodic lattice. The Bloch theorem²² states that a single-electron wavefunction $\phi_{\alpha\mathbf{k}}$ is composed of a plane wave with wave vector \mathbf{k} and a lattice periodic part $u_{\alpha\mathbf{k}}$:

$$\phi_{\alpha\mathbf{k}}(\mathbf{r}) = u_{\alpha\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (1)$$

The plane-wave part describes the effective motion of electrons throughout the crystal, whereas $u_{\alpha\mathbf{k}}(\mathbf{r})$ encodes the contribution from atomic orbitals and spins (quantum number α). In conventional electronics, quasi-static fields cause a tiny imbalance of the number of electrons moving in the $+\mathbf{k}$ and $-\mathbf{k}$ directions. This leads to incoherent charge currents. In quantum electronics, in contrast, one tries to coherently transport the quantum information stored in atomic orbitals or spins before coherence-destroying scattering becomes effective²³. Recently, a carrier wave of intense light pulses has been exploited as an alternating-current bias to transport electrons faster than scattering occurs, yielding intriguing quantum effects, such as dynamical Bloch oscillations^{5,7,11,13}, electron–hole recollisions^{6,12,24}, electronic quantum interference^{8,9}, and interband excitations^{8–10}. The resulting high-harmonic radiation^{5,7,8,10–15} has been used to retrieve the electronic band structure²⁵ optically. Despite the first evidence of Berry's phase effects^{13,26}, resulting directly from the topological \mathbf{k} dependence of $u_{\alpha\mathbf{k}}$, most lightwave-driven transport of electric charge has focused on controlling \mathbf{k} in the plane-wave part of the Bloch wavefunction.

Independently, quantum information processing² has been pursued using internal degrees of freedom encoded in $u_{\alpha\mathbf{k}}$, such as spin¹⁶ or valley pseudospin^{17–21,27–30}. Monolayers of transition-metal dichalcogenides form an ideal testbed for such concepts because they feature

two separate band minima—called valleys^{17–19}—that can be selectively excited by optical means^{18,19,27–29}. The coherence of the valley pseudospin has been shown to be conserved even in steady-state photoluminescence experiments^{27–29}, rendering this degree of freedom a promising information carrier^{17–19,21,27–30}. Yet its subcycle manipulation remains an open challenge.

Here we demonstrate that the valley pseudospin in a transition-metal dichalcogenide monolayer can be changed by lightwave-driven intra-band transport within a few femtoseconds. To this end, we accelerate optically prepared coherent electron–hole pairs in monolayer tungsten diselenide (WSe₂) by intense, phase-stable multi-terahertz waveforms¹². Recolliding electrons and holes emit their kinetic energy in high-order sidebands, the odd orders of which directly reflect the electron–hole pairs' valley pseudospin. Pairs that are selectively created in a single valley can be transferred partly into the opposite valley by the strong terahertz field, leaving a unique fingerprint in elliptically polarized sidebands.

Monolayers of WSe₂ (Fig. 1a) feature an electronic band structure with a direct energy gap separating the filled valence band from the empty conduction band at the corners of the hexagonal unit cell in momentum space, called the Brillouin zone^{17–19} (Fig. 1b). Owing to the honeycomb crystal lattice (Fig. 1a), the Brillouin zone contains two inequivalent corners^{17–19}—the K and the K' points. In the vicinity of K (or K'), the electronic wavefunction $u_{\alpha\mathbf{k}}$ of the highest valence band is primarily composed of atomic 5d orbitals of tungsten with a magnetic quantum number^{18,19} $m = 2$ (or -2). A sign flip of m implies a time inversion of the electronic wavefunction $u_{\alpha\mathbf{k}}$ between K and K' (insets in Fig. 1b). Also the conduction-band wavefunctions at the K and K' points are time-reversal pairs. This is why circularly polarized light can prepare electron–hole pairs selectively at the K or the K' point, depending on helicity^{17–21,27–29}, and the valley polarization may be described by a spin-like quantity called valley pseudospin. In the following, we determine whether intense terahertz pulses can drive electron transport far enough to induce superpositions of K and K' states and ultimately to flip the valley pseudospin.

We start with a linearly polarized 100-fs near-infrared pulse to create electron–hole pairs in bulk and monolayer WSe₂ on a diamond substrate (Extended Data Fig. 1). Being a superposition of right- (σ_+) and left-circularly (σ_-) polarized light, this preparation pulse generates coherent electron–hole pairs in both the K and K' valleys. Coulomb attraction causes the quasiparticles to form a series of atom-like bound states, called excitons²³. Our preparation pulse is resonant with the 1s A series (transition from the uppermost valence band to the conduction band) exciton^{12,17–21} (with photon energies 1.621 eV for the bulk and 1.665 eV for the monolayer). Simultaneously, an intense terahertz transient (inset in Fig. 1c; centre frequency, $\nu_{\text{THz}} = 40$ THz) coherently drives the excitonic polarization, leading to real-space acceleration followed by a recollision of the constituent electron and hole. This dynamics is reminiscent of the three-step model underlying attosecond pulse generation in atomic

¹Department of Physics, University of Regensburg, Regensburg, Germany. ²Department of Physics, University of Marburg, Marburg, Germany. ³Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. ⁴These authors contributed equally: F. Langer, P. G. Hawkins. *e-mail: mackkira@umich.edu

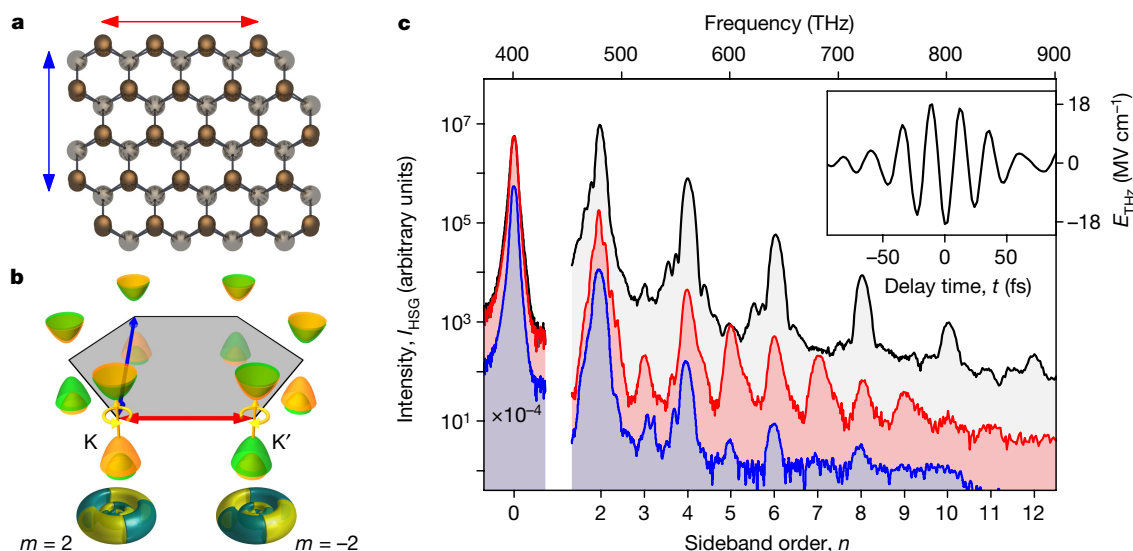


Fig. 1 | High-odd-order sideband generation in monolayer WSe₂. **a**, Crystal structure of a WSe₂ monolayer. Tungsten atoms are depicted as grey, selenium atoms as brown spheres. **b**, Hexagonal Brillouin zone with the spin-split valence and conduction band edges at the inequivalent K and K' points. The main orbital contributions to the top of the valence bands, as calculated from density functional theory, are shown below (the green–yellow colour scale depicts the sign of the imaginary part). Red (blue) arrows in **a** and **b** indicate the zigzag (armchair) direction. **c**, High-order sideband intensity, I_{HSG} , measured from bulk WSe₂ (black)

and monolayer WSe₂ in the zigzag (red) and the armchair direction (blue, shifted downwards by one order of magnitude for a better visibility). In the zigzag direction, even- and odd-order sidebands exhibit comparable intensities, while odd orders are almost completely suppressed in the armchair direction. We note that the frequency axis of the bulk spectrum has been shifted by 9 THz owing to the different excitation frequencies. The inset shows the electro-optically sampled multi-terahertz field, featuring a frequency of $\nu_{\text{THz}} = 40$ THz and a peak field strength in air of $E_{\text{THz}} = 18$ MV cm⁻¹.

gases⁴. Upon recollision, the electron can recombine with the hole, emitting a photon with an energy corresponding to the bandgap plus the kinetic energy imparted by the acceleration.

In bulk WSe₂, this mechanism leads to the generation of high-order sidebands accompanying the interband 1s exciton line (Fig. 1c, black curve). Because the bulk crystal is inversion-symmetric, high-order sideband generation (HSG) is independent of the polarity of the driving field, and collisions between the excitonic constituents occur in every half-cycle. In the frequency domain, this periodicity causes even sideband orders with a spectral separation of $2\nu_{\text{THz}}$.

Figure 1c also depicts an observation of high-order sideband emission from a monolayer material (red curve), the first to our knowledge. Despite the ultimately low sample thickness, harmonic sidebands up to order $n = 11$ clearly rise above the noise floor. Remarkably, the sideband spectrum differs qualitatively from the bulk case. Besides even orders, we also detect odd orders ($n > 3$) of comparable strength—a result that has neither been predicted nor observed before. Phenomenologically, the emergence of HSG spaced by ν_{THz} implies that the time structure of the light emission is periodic after a full oscillation cycle of the terahertz wave, that is, it differs for positive and negative terahertz half-cycles.

For a microscopic understanding, we first note that the strongest odd-order sideband intensity occurs when electron–hole pairs are accelerated along the zigzag crystal direction (Fig. 1c, red curve), whereas HSG is strongly suppressed for terahertz fields along the armchair direction (Fig. 1c, blue curve). Because inversion symmetry along the zigzag direction is broken only in momentum space (Fig. 1b, red arrow), not in real space (Fig. 1a, red arrow), our observation proves that odd-order sideband generation must be related to intraband transport, which is best understood in momentum space. The terahertz field drives electrons and holes within the bands in which the near-infrared pulse has prepared them. The emergence of odd-order HSG indicates that the coherent electron–hole pairs traverse a large fraction of the Brillouin zone to reach wave vectors \mathbf{k} at which the effective mass is no longer isotropic. Hence, the carrier dynamics depends on the sign of the driving waveform and a spectral modulation of the sideband radiation with ν_{THz} results.

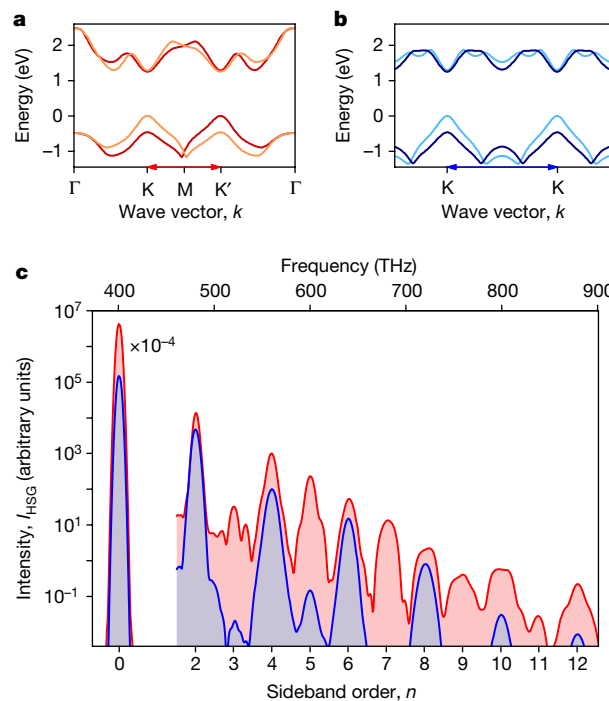


Fig. 2 | Electronic structure of WSe₂ and quantum theory of high-order sideband generation. **a**, **b**, Band structure of monolayer WSe₂ calculated by DFT for the zigzag (**a**) and the armchair (**b**) directions (dark colours, spin-up bands; light colours, spin-down bands). There is no inversion symmetry about the K/K' valleys in the zigzag direction, supporting the generation of odd-order sidebands. The bands in the armchair direction are inversion-symmetric, inhibiting odd-order HSG. The red and blue arrows at the bottom of each panel indicate the high-symmetry directions highlighted in Fig. 1. **c**, Computed intensity spectrum I_{HSG} for lightwave-driven electron–hole recollision along the zigzag (red) and armchair (blue, down-shifted along the intensity axis by one order of magnitude for better visibility) directions.

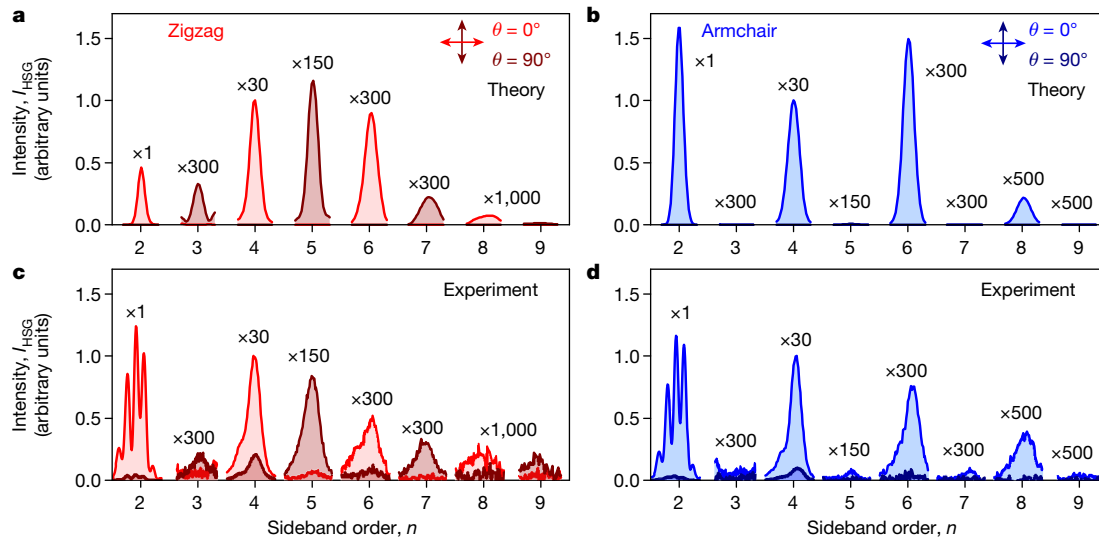


Fig. 3 | High-order sideband polarization for different crystal orientations. **a, b**, Simulated high-order sideband intensity in the polarization basis parallel ($\theta = 0^\circ$, light colours) and perpendicular ($\theta = 90^\circ$, dark colours) to the linearly polarized excitation and driving fields, taking into account the polarization contributions of opposite helicity in the different valleys. Individual sideband orders are multiplied with the indicated multiplication factors. Even and odd orders are cross-polarized in the zigzag direction (**a**), while the armchair direction (**b**)

supports only strong even orders with a parallel polarization. **c, d**, Measured polarization-resolved intensity I_{HSG} normalized to the fourth order for a multi-terahertz field pointing along the zigzag (**c**) and the armchair (**d**) directions. The zigzag direction confirms cross-polarized even- and odd-order sidebands, while all orders in the armchair direction exhibit a parallel polarization (odd orders suppressed by crystal symmetry).

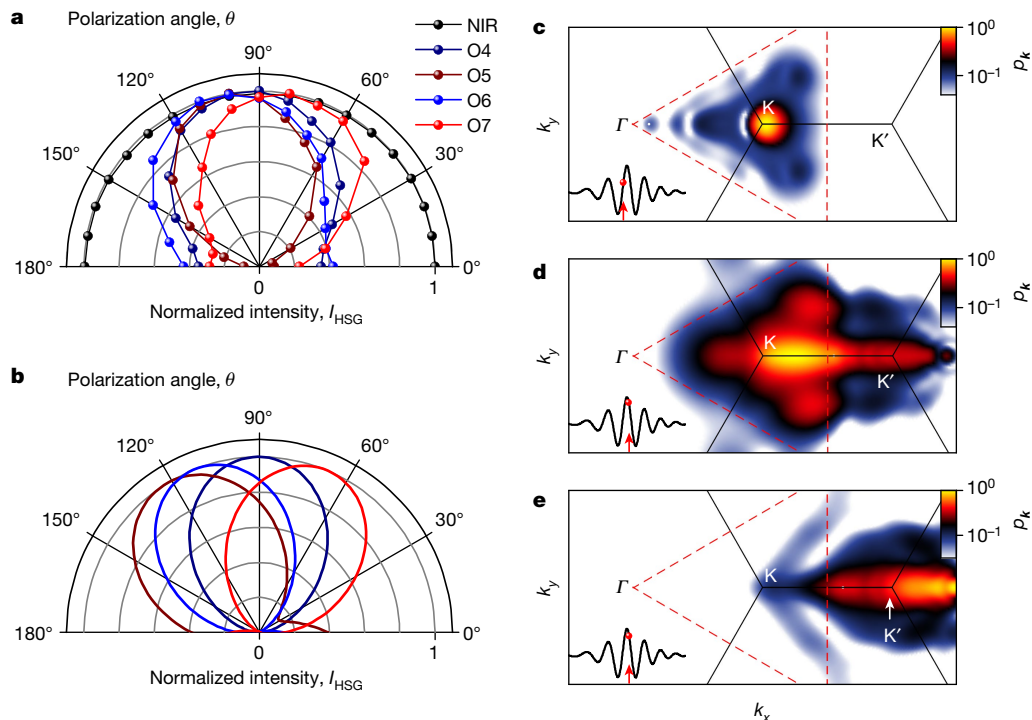


Fig. 4 | Intervalley mixing and lightwave valleytronics. **a**, Measured polarization-resolved intensity I_{HSG} of individual sideband orders (normalized, coloured data points) for circularly polarized excitation (normalized, black data points) of the K valley only. The terahertz-driven intraband transport ($E_{\text{THz}} = 23 \text{ MV cm}^{-1}$, along the zigzag direction) into the K' valley adds polarization contributions of opposite helicity, resulting in an overall elliptical polarization. **b**, Computed polarization of high-order sidebands (orders four to seven, solid lines, same colours as in **a**) following a valley-selective excitation in the K valley. **c, d**, Computed distribution of the coherent electron-hole polarization p_k (colour scale) in reciprocal space at the time of small (**c**) and maximal displacement (**d**);

conditions match the experimental analysis in **a**. Near a zero-crossing of the electric field (indicated in the inset of **c**), the coherent electron-hole pairs reside in the K valley (borders indicated by red dashed lines). In the following half-cycle (**d**), a pronounced ledge of electrons and holes is driven from the K to the K' valley, underpinning subcycle valley transport. **e**, Computed polarization following a subcycle excitation by a 5-fs optical pulse and terahertz acceleration by an electric field of 23 MV cm^{-1} . The maximum of the polarization is transferred far into the K' valley, demonstrating that the valley pseudospin can indeed be switched by strong terahertz fields.

We corroborate this conjecture by computing the full band structure of the WSe₂ monolayer with density functional theory (DFT; see Methods). For large excursions from the K or K' points, the electron–hole pairs do indeed experience strong asymmetries of the band structure along the zigzag direction (Fig. 2a). In contrast, the band structure in the armchair direction is symmetric about the K and the K' points (Fig. 2b). Consequently, the sideband spectrum is expected to show hardly any odd orders, as indeed is observed in Fig. 1c.

Nevertheless, this explanation is incomplete because the linearly polarized preparation pulse populates both K and K' valleys equally. Since the bands in these valleys are mirror images of each other (see Fig. 2a), the plane-wave part of the electron–hole dynamics of the ensemble is effectively symmetric for positive and negative terahertz half-cycles, which forbids the generation of odd-order sidebands. This obvious contradiction with the experiment can be resolved only if the valley pseudospin, encoded in the atomic wavefunction $u_{\alpha k}$, is taken into account. Therefore, we simulate the entire life cycle of the electron–hole pairs by combining a DFT calculation of the electronic band structure and the Bloch states with cluster expansion²³ to solve the lightwave-driven non-perturbative many-body dynamics. We quantitatively include the electron–hole Coulomb interaction, pseudospin and dipole-matrix elements in two-dimensional computations for conditions matching the experiment (see Methods).

The computed high-order sideband spectra (Fig. 2c) reproduce the experimental results very well, including the emergence of even and odd orders as well as details, such as the suppression of the third-order sideband (see Methods). Most importantly, the presence of odd-order sidebands is observable in our setting only if the valley pseudospin is accounted for. The different pseudospin in the K and K' valleys imparts different helicities on the sideband radiation resulting from recombination within the respective valleys. This fact breaks the symmetry for positive and negative terahertz field polarities and enables the generation of odd orders. Interestingly, the pseudospin tagging of the K and K' excitations causes a characteristic polarization state of the sidebands (Fig. 3a and b). Straightforward vector algebra (see Methods) shows that summing up the counter-circularly rotating field vectors of odd-order sidebands from the K and K' valleys yields a field that is linearly polarized perpendicular to the excitation pulse. Analogously, one can show that even-order sidebands have to be polarized parallel with the excitation light.

We test this prediction by measuring the intensity of even and odd sideband orders with an analyser set at angle θ (Fig. 3c and d). Here $\theta = 0^\circ$ (or 90°) denotes a polarization parallel (or perpendicular) to the near-infrared and the terahertz fields. As expected in the zigzag direction, even-order sidebands are indeed polarized parallel to the excitation light, while odd orders are perpendicularly polarized (Fig. 3a and c). By crystal symmetry, the armchair direction supports only even orders, which are polarized at $\theta = 0^\circ$ (Fig. 3b and d). This theory–experiment comparison further corroborates that odd-order HSG is a fingerprint of the valley coherence, which prevails even under atomically strong fields.

To trace precisely how the terahertz field changes the valley pseudospin, we selectively excite the K valley by a σ_+ polarized excitation pulse. If the electron–hole pairs recollide within the same valley, their sideband emission is expected to remain polarized with the same helicity. Figure 4a shows the σ_+ polarization of the near-infrared preparation pulse (black curve), as well as the polarization state of the sideband orders $n = 4$ to 7, recorded by measuring the intensity as a function of the angle of a rotating analyser. In stark contrast to the incident pulse, the sidebands are polarized strongly elliptically, with their principal axis aligned at a large angle θ (Fig. 4a). Our many-body computations (Fig. 4b) produce essentially the same polarization and quantitatively connect it with coherent K-to-K' valley transport, which moves the electron–hole pairs into a superposition of both valleys contributing opposite helicities to the sideband emission.

The ellipticity of the resulting polarization is a measure of the yield of the intervalley transfer whereas θ is set by the relative phase of the wave functions in the K and the K' valley. Both dynamical and geometric phases contribute. In the future, the delicate variations in the polarization direction of different orders may be quantitatively evaluated to extract geometric phases acquired during lightwave acceleration within the bands.

To visualize the intervalley transport, we calculate the full two-dimensional distribution of the coherent electron–hole polarization p_k in reciprocal space, which is proportional to the density of coherent electron–hole pairs (see Methods). At a zero-crossing of the field, the electrons and holes reside in the K valley (Fig. 4c) while the following half-cycle accelerates the leading edge of p_k into the K' valley within a few femtoseconds (Fig. 4d), attesting to subcycle coherent intervalley transport. The distribution function shows major spreading and deformation from the single-peak excitonic packet, mainly due to ionization-induced deformation of the wave packet, whereas a peak close to K is caused by the continued optical excitation during the terahertz acceleration.

A key question is how far intervalley transport may be ultimately driven. For a realistic scenario, we compute the distribution of p_k for a terahertz peak field of 23 MV cm⁻¹ and a shorter 5-fs near-infrared pulse to prepare electron–hole pairs at a well defined phase of the terahertz carrier wave. Figure 4e shows the coherent distribution at its maximal displacement, when the terahertz field has passed an intense half-cycle. Almost the complete distribution of coherent electron–hole pairs (96%) is driven into the K' valley. Here the time evolution of the quantum mechanical wavefunction is effectively inverted by adiabatically shifting it through a time-reversal-invariant momentum point. These calculations are supported by a first experiment that utilizes subcycle injection with a 10-fs optical pulse and transfers 66% of the coherent electron–hole pairs into the K' valley (see Extended Data Fig. 2). We anticipate that custom-tailored terahertz waveforms and injection times could allow for yet more sophisticated transfer protocols, paving the way to ultimately fast valleytronics.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0013-6>.

Received: 21 September 2017; Accepted: 16 February 2018.

Published online 2 May 2018.

- Markov, I. L. Limits on fundamental limits to computation. *Nature* **512**, 147–154 (2014).
- Ladd, T. D. et al. Quantum computers. *Nature* **464**, 45–53 (2010).
- Krausz, F. & Stockman, M. I. Attosecond metrology: from electron capture to future signal processing. *Nat. Photon.* **8**, 205–213 (2014).
- Corkum, P. B. & Krausz, F. Attosecond science. *Nat. Phys.* **3**, 381–387 (2007).
- Ghimire, S. et al. Observation of high-order harmonic generation in a bulk crystal. *Nat. Phys.* **7**, 138–141 (2011).
- Zaks, B., Liu, R. B. & Sherwin, M. S. Experimental observation of electron–hole recollisions. *Nature* **483**, 580–583 (2012).
- Schubert, O. et al. Sub-cycle control of terahertz high-harmonic generation by dynamical Bloch oscillations. *Nat. Photon.* **8**, 119–123 (2014).
- Hohenleutner, M. et al. Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).
- Higuchi, T., Heide, C., Ullman, K., Weber, H. B. & Hommelhoff, P. Light-field-driven currents in graphene. *Nature* **550**, 224–228 (2017).
- Vampa, G. et al. Linking high-harmonics from gases and solids. *Nature* **522**, 462–464 (2015).
- Garg, M. et al. Multi-petahertz electronic metrology. *Nature* **538**, 359–363 (2016).
- Langer, F. et al. Lightwave-driven quasiparticle collisions on a subcycle timescale. *Nature* **533**, 225–229 (2016).
- Liu, H. et al. High-harmonic generation from an atomically thin semiconductor. *Nat. Phys.* **13**, 262–265 (2017).
- Yoshikawa, N., Tamaya, T. & Tanaka, K. High-harmonic generation in graphene enhanced by elliptically polarized light excitation. *Science* **356**, 736–738 (2017).
- Sivis, M. et al. Tailored semiconductors for high-harmonic optoelectronics. *Science* **357**, 303–306 (2017).

16. Wolf, S. A. et al. Spintronics: a spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
17. Schaibley, J. R. et al. Valleytronics in 2D materials. *Nat. Rev. Mater.* **1**, 16055 (2016).
18. Xu, X., Wang, Y., Xiao, D. & Heinz, T. F. Spin and pseudospins in layered transition metal dichalcogenides. *Nat. Phys.* **10**, 343–350 (2014).
19. Xiao, D., Liu, G.-B., Feng, W., Xu, X. & Yao, W. Coupled spin and valley physics in monolayers of MoS₂ and other group-VI dichalcogenides. *Phys. Rev. Lett.* **108**, 196802 (2012).
20. Aivazian, G. et al. Magnetic control of valley pseudospin in monolayer WSe₂. *Nat. Phys.* **11**, 148–152 (2015).
21. Ye, Z., Sun, D. & Heinz, T. F. Optical manipulation of valley pseudospin. *Nat. Phys.* **13**, 26–29 (2017).
22. Bloch, F. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. Phys.* **52**, 555–600 (1929).
23. Kira, M. & Koch, S. W. *Semiconductor Quantum Optics* (Cambridge Univ. Press, Cambridge, 2012).
24. Yan, J.-Y. Theory of excitonic high-order sideband generation in semiconductors under a strong terahertz field. *Phys. Rev. B* **78**, 075204 (2008).
25. Vampa, G. et al. All-optical reconstruction of crystal band structure. *Phys. Rev. Lett.* **115**, 193603 (2015).
26. Banks, H. B. et al. Dynamical birefringence: electron–hole recollisions as probes of Berry curvature. *Phys. Rev. X* **7**, 041042 (2017).
27. Mak, K. F., He, K., Shan, J. & Heinz, T. F. Control of valley polarization in monolayer MoS₂ by optical helicity. *Nat. Nanotechnol.* **7**, 494–498 (2012).
28. Jones, A. M. et al. Optical generation of excitonic valley coherence in monolayer WSe₂. *Nat. Nanotechnol.* **8**, 634–638 (2013).
29. Wang, G. et al. Control of exciton valley coherence in transition metal dichalcogenide monolayers. *Phys. Rev. Lett.* **117**, 187401 (2016).
30. Rycerz, A., Tworzydło, J. & Beenakker, C. W. J. Valley filter and valley valve in graphene. *Nat. Phys.* **3**, 172–175 (2007).

Acknowledgements The work in Regensburg was supported by the European Research Council through grant number 305003 (QUANTUMsubCYCLE) as well as by the Deutsche Forschungsgemeinschaft (through grant number HU 1598/2-1, SFB 1277, projects A05, B05 and B06, and GRK 1570) and the work in Marburg and Michigan by the Deutsche Forschungsgemeinschaft (through SFB 1083 and grant numbers KI 917/3-1 and KI 917/2-2).

Reviewer information *Nature* thanks J. Wang and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions F.L., P.G.H., C.P.S., S.S., S.W.K., M.K. and R.H. conceived the study. F.L., C.P.S., S.S. and R.H. carried out the experiment and analysed the data. P.N., C.S. and T.K. provided, processed and characterized the samples. M.G. and J.F. performed the DFT calculations and P.G.H., J.T.S., U.H., S.W.K. and M.K. developed the quantum-mechanical model, carried out the computations and analysed the data. All authors discussed the results and contributed to the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41586-018-0013-6>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to M.K.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Experimental set-up. A femtosecond titanium-sapphire laser amplifier (repetition rate, 3 kHz; pulse energy, 5.5 mJ; pulse duration, 33 fs; centre wavelength, 805 nm) pumps two parallel dual-stage optical parametric amplifiers delivering signal pulses with energies of up to 0.5 mJ and centre wavelengths that are tuneable between 1.1 μm and 1.6 μm . We generate intense, phase-locked waveforms in the far- to mid-infrared spectral region (multi-terahertz range) via difference-frequency generation between the spectrally detuned phase-correlated near-infrared pulse trains of the optical parametric amplifier. The electric peak field of these few-cycle pulses reaches values of up to $1 \text{ V } \text{\AA}^{-1}$. A yttrium aluminium garnet (YAG)-based super-continuum source provides ultrabroadband excitation pulses, covering the near-infrared and visible spectral ranges. The excitation pulse and the terahertz driving field are superimposed with an indium-tin-oxide-coated beam splitter and are collinearly focused onto a tungsten diselenide sample by a gold-coated parabolic mirror. The full-width-at-half-maximum (FWHM) spot sizes of the excitation pulse and the terahertz transient at the sample position are 22 μm and 50 μm (FWHM of the intensity), respectively. For our experiments, the centre frequency of the terahertz driving field is set to $\nu_{\text{THz}} = 40 \text{ THz}$, while the peak electric field in air amounts to $E_{\text{THz}} = 18 \text{ MV cm}^{-1}$ unless stated otherwise. With the help of a mechanical delay stage in the excitation beam path, the temporal overlap of both pulses is adjusted for maximum sideband emission. For a resonant excitation of the excitonic 1s ground state (with photon energies 1.621 eV for the bulk and 1.665 eV for the monolayer) in WSe_2 kept at room temperature, optical band-pass filters with a bandwidth of 10 nm are employed to spectrally constrain the white-light pulses. The generated high-order sideband radiation is recorded with a spectrograph coupled to a thermoelectrically cooled silicon charge-coupled device (CCD) camera. All spectra are corrected for the diffraction efficiency of the grating and the quantum efficiency of the detector, as well as the sensitivity of the spectrograph, for different polarizations. For the valley-selective injection of electron-hole pairs with σ_+ polarized light, we employ a combination of a half-wave and a quarter-wave plate in the beam path of the white-light pulses. The polarization states of the excitation pulses and the emitted sidebands are analysed by a Glan-Thompson polarizer (extinction ratio, 10^{-5}) placed after the sample. Employing ultrashort gating pulses (pulse duration, 10 fs), we detect the waveform of the terahertz transients by electro-optic sampling in a 6.5- μm -thick zinc telluride crystal while accounting for the detector response³¹.

Sample preparation. Bulk-like and monolayer samples of tungsten diselenide are mechanically exfoliated from a bulk crystal onto an intermediate substrate, a visco-elastic gel film (polydimethylsiloxane; Extended Data Fig. 1b). There, the sample thickness is verified using optical contrast. Subsequently, the samples are transferred to a diamond substrate grown by chemical vapour deposition (Extended Data Fig. 1c). In this dielectric environment, the exciton binding energy³² amounts to 0.25 eV, corresponding to a Keldysh parameter of 0.16 for the field ionization of the exciton using a terahertz amplitude of 18 MV cm^{-1} . The energy of the 1s A-exciton state is confirmed by absorption measurements employing the white-light pulses, as well as photoluminescence emission under continuous-wave excitation of a green laser diode operating at a wavelength of 533 nm. All experiments are performed under ambient conditions. The crystal orientation is determined by second-harmonic generation. For this purpose, linearly polarized, 10-fs near-infrared pulses are focused onto the monolayer sample. The second-harmonic intensity with a polarization parallel to the fundamental, $I_{\text{SHG},||}$, is monitored using a spectrograph with a cooled silicon CCD. $I_{\text{SHG},||}$ peaks in the armchair direction (Extended Data Fig. 1a), where inversion symmetry is explicitly broken in the real space lattice (compare to Fig. 1a). Because monolayer WSe_2 preferably cleaves along high-symmetry directions, the alignment procedure is complemented by optical microscope images (Extended Data Fig. 1b, c). We estimate that the crystal orientation is determined with an uncertainty of approximately $\pm 1^\circ$.

DFT calculations. Our DFT calculations are performed with the full-potential linearized augmented plane-wave all-electron method, as implemented in the Wien2k code³³. Specific calculation parameters and an optimized structure of WSe_2 were taken from Kormányos et al.³⁴. The wavefunction (with muffin-tin radii of 2.46 and 2.34 atomic units for W and Se, respectively) was expanded in partial waves with orbital quantum numbers up to 10. For the interstitial region, we used a plane-wave cut-off of 6.5 \AA^{-1} . The spin-orbit interaction was included fully relativistically for core electrons while $W 5s^2 5p^6 4f^{14} 5d^4 6s^2$ and $Se 3d^{10} 4s^2 4p^4$ valence electrons were treated within a second variational step method³⁵. For the exchange-correlation functional, we consider the generalized gradient approximation³⁶. These DFT calculations provide the relevant electronic bands and dipole matrix elements (including both σ_+ and σ_- contributions) throughout the irreducible Brillouin zone wedge.

Because DFT solves electronic states and matrix elements from the atomistic level up, by using minimal assumptions, it is a first-principles approach. Nevertheless, DFT must always approximate the correlation energy functional

because its dependence on density still remains unknown. To reproduce the experimental bandgap and exciton resonance energies, we adjust both the bandgap and dielectric constant of the Coulomb interaction V . We apply the Keldysh form for V to account for the principal aspects of dielectric screening inside a monolayer. These inputs uniquely define both the Coulomb and light-matter interactions needed to systematically describe electron-hole excitations generated by an optical and a terahertz field. These procedures yield an accurate description of the excitonic features of transition-metal dichalcogenides³⁷.

Cluster-expansion computations. The cluster-expansion approach solves the many-body problem from the point of view of correlations, as an alternative perspective compared to DFT. Earlier investigations with semiconductors²³, the Jaynes-Cummings model³⁸ and strongly interacting atomic Bose-Einstein condensates³⁹ show that many-body dynamics can be cast into a format exactly described in terms of a sequential build-up of clusters. Because cluster expansion defines all the resulting correlation functionals exactly⁴⁰ for any given many-body Hamiltonian, it is also a first-principles approach. The description, furthermore, is non-perturbative in terms of the interaction strength^{41,42}, which makes it ideally suited for studying extreme nonlinearities created, for example, by HSG. Because clusters build up sequentially, the dynamics on extremely short timescales is well described by completely omitting clusters beyond a certain level of complexity. This circumstance makes the cluster-expansion computations extremely efficient and accurate in describing ultrafast quantum kinetics of very diverse many-body systems.

Here, the optical field $E_{\text{opt}}(t)$ predominantly excites a coherent superposition between the valence and the conduction band, that is, a microscopic polarization p_k . The polarization p_k can be characterized by a single-particle cluster^{23,41} describing coherently driven electron-hole pairs at crystal momentum $\hbar k$. The polarization dynamics forms the backbone of the semiconductor Bloch equations; its polarization part has the structure

$$i\hbar \frac{\partial}{\partial t} p_k = \epsilon_k p_k - (1 - f_k^e - f_k^h) \Omega_k(t) + i|e| E_{\text{THz}}(t) \cdot \nabla_k p_k + \Gamma_k$$

where ϵ_k is the Coulomb-renormalized electron-hole energy (defined by the DFT input), and f_k^e (f_k^h) is the electron (hole) occupation at k , $\Omega_k(t) = d_k \cdot E_{\text{opt}}(t) + \sum_{k'} V_{k'-k} p_{k'}$ is the Rabi energy renormalized by the Coulomb interaction V_k . Both V_k and the dipole-matrix element d_k are defined from DFT. Once the optical field $E_{\text{opt}}(t)$ generates the polarization, the terahertz field $E_{\text{THz}}(t)$ accelerates it via the gradient term. The generated microscopic polarization also scatters with densities and other polarization p_k due to Coulomb interaction, inducing a two-particle correlation Γ_k that introduces exciton-state-dependent dephasing for polarization. This full microscopic scattering is included using the description of Smith et al.⁴³. This yields a fully microscopic description of exciton-state-dependent dephasing, which considerably decreases the coherence time for all excitons except the 1s state⁴¹. This level of sophistication is necessary to systematically model the decay of coherent electron-hole pairs during their life cycle, encompassing preparation, acceleration and recollision. Because transition-metal dichalcogenide monolayers couple strongly to light, we also include the self-consistent coupling of the semiconductor Bloch equations with Maxwell's equations, following the exact approach of Kira and Koch⁴¹, using the experimental geometry vacuum-monolayer-diamond-vacuum. This allows us to predict both the transmitted and reflected fields in absolute units, as well as to include the radiative decay in the analysis. We use the DFT material parameters and matrix elements as an input to solve the full two-dimensional quantum polarization kinetics using 170 radial wave vectors spanning the full Brillouin zone and 241 angular states.

HSG and symmetry. Broken inversion symmetry is a necessary requirement for odd-order sideband generation. Inversion-symmetric media, such as bulk crystals of silicon or diamond, do not facilitate odd-order HSG. Broken out-of-plane symmetry at surfaces, interfaces and two-dimensional materials can manifest itself only if the driving field has a strong out-of-plane component. As seen in Fig. 1c, the strongest effect of HSG occurs in crystal directions in which the symmetry of the band structure, rather than the symmetry of the real-space lattice, is broken. This is caused by the fact that HSG originates from intraband currents. Interestingly, the third-order sideband in Fig. 1c is relatively strongly suppressed. This feature is indeed expected (Fig. 2) because this sideband results from electron-hole excursions that barely reach the asymmetric band structure features, whereas higher-order sidebands are associated with substantially longer trajectories in k space. Nonetheless, even in cases where the band structure is asymmetric, odd-order HSG may require an additional condition. Our experiment with linearly polarized excitation light exemplifies the situation. Since the band structure is mirror-symmetric with respect to the M point, the intraband dynamics of electron-hole pairs in the K valley driven by a positive terahertz field is identical to the intraband dynamics of electron-hole pairs in the K' valley driven by a sign-inverted

terahertz field. For equal populations in both valleys, the total system may seem inversion-symmetric at first glance. The observation of odd-order HSG (Fig. 1c, red curve) can be explained only if the valley pseudospin is taken into account: electron-hole recollisions from different valleys are disentangled by the valley-specific helicity of the emitted light. Alternatively, the contributions of both valleys can be separated by exciting only one valley with circularly polarized light (Fig. 4a). In any case, the emergence of odd-order sidebands in monolayer WSe₂ is a direct consequence of the valley pseudospin in the material. We further underpin the influence of the valley pseudospin and intervalley transfer by a first experiment using subcycle injection of coherent electron-hole pairs by a 10-fs pulse (Extended Data Fig. 2). We extract a transfer rate of 66% marking the first subcycle transport of internal quantum attributes, which could inspire a development of corresponding quantum logic operations similar to spintronics⁴⁴.

Analytical derivation of the high-order sideband polarization. The theoretical and experimental finding of cross-polarized even and odd sideband orders in the case of a linearly polarized interband excitation (see Fig. 3) is a direct consequence of the opposite pseudospins of the K and K' valleys and can be derived analytically as follows. Owing to the optical selection rules of monolayer WSe₂, radiative electron-hole recombination in the two valleys generates light of opposite helicity. Consequently, the total emitted field of high-order sidebands E_{HSG} consists of contributions from the K and K' valley with right- (σ_+) and left-circularly polarized (σ_-) light, respectively, and can be written as

$$E_{\text{HSG}} = E_{\text{HSG,K}}\sigma_+ + E_{\text{HSG,K'}}\sigma_-$$

where $E_{\text{HSG,K}}$ and $E_{\text{HSG,K'}}$ denote the contributions from the K and the K' valley, respectively. Owing to the symmetry of the band structure (see Fig. 2a), the dynamics occurring in the K' valley for positive fields of the driving waveform E_{THz} is identical to the dynamics of the K valley for negative fields (K and K' points are time-reversal pairs). Hence, we can substitute $E_{\text{HSG,K'}}(E_{\text{THz}})$ with $E_{\text{HSG,K}}(-E_{\text{THz}})$. By simultaneously using the relation $\sigma_{\pm} = (\mathbf{e}_x \mp i\mathbf{e}_y)/\sqrt{2}$, with the Cartesian unit vectors \mathbf{e}_x and \mathbf{e}_y , we can rewrite the above relation as

$$E_{\text{HSG}} = E_{\text{HSG,K}}(E_{\text{THz}}) (\mathbf{e}_x - i\mathbf{e}_y)/\sqrt{2} + E_{\text{HSG,K}}(-E_{\text{THz}}) (\mathbf{e}_x + i\mathbf{e}_y)/\sqrt{2}$$

Furthermore, one can decompose $E_{\text{HSG,K}}$ into even- and odd-order contributions, that is, $E_{\text{HSG,K}} = E_{\text{even}} + E_{\text{odd}}$. Because only even (odd) powers of E_{THz} contribute to E_{even} (E_{odd}), these fields feature even (odd) parity with respect to the terahertz driving field. Taking these relations into account, one can state

$$E_{\text{HSG,K}}(\pm E_{\text{THz}}) = E_{\text{even}} \pm E_{\text{odd}}$$

and we obtain

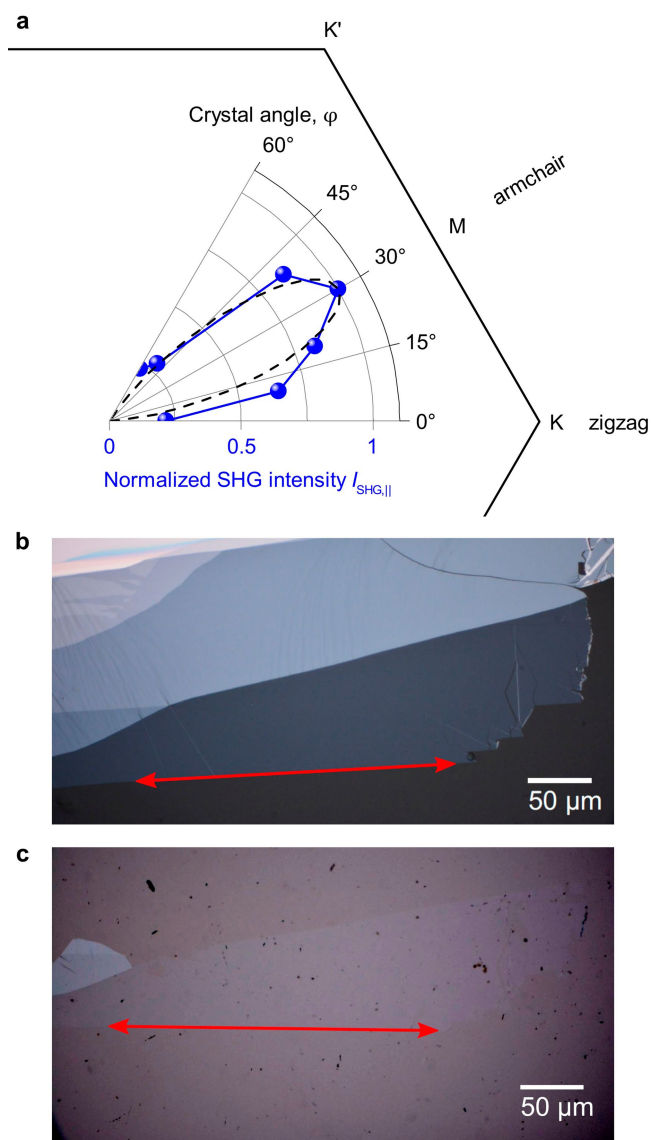
$$E_{\text{HSG}} = (E_{\text{even}} + E_{\text{odd}}) (\mathbf{e}_x - i\mathbf{e}_y)/\sqrt{2} + (E_{\text{even}} - E_{\text{odd}}) (\mathbf{e}_x + i\mathbf{e}_y)/\sqrt{2}$$

$$E_{\text{HSG}} \propto E_{\text{even}} \cdot \mathbf{e}_x - E_{\text{odd}} \cdot i\mathbf{e}_y$$

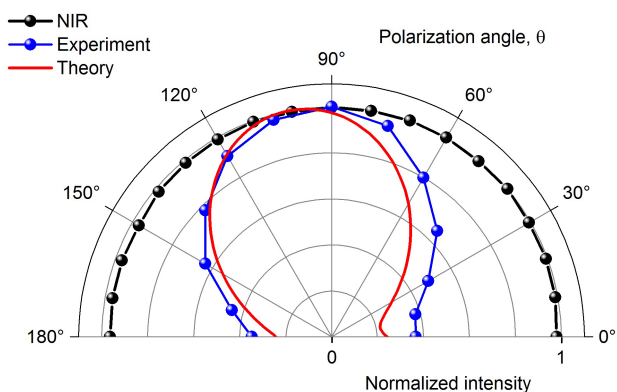
Consequently, the conserved circular dichroism directly leads to cross-polarized even and odd orders. These derivations still hold for large intervalley transfer as long as the optical selection rules hold. Our experimental observations confirm this fact as well as the conservation of the valley coherence during lightwave acceleration.

Data availability. The data that support the findings of this study are available from the corresponding author upon request.

31. Gallot, G. & Grischkowsky, D. Electro-optic detection of terahertz radiation. *J. Opt. Soc. Am. B* **16**, 1204–1212 (1999).
32. Poellmann, C. et al. Resonant internal quantum transitions and femtosecond radiative decay of excitons in monolayer WSe₂. *Nat. Mater.* **14**, 889–893 (2015).
33. Blaha, P., Schwarz, K., Madsen, G. K. H., Kvasnicka, D. & Luitz, J. *Wien2k, An Augmented Plane Wave+Local Orbitals Program for Calculating Crystal Properties* <http://susi.theochem.tuwien.ac.at/> (Vienna Univ. Technology, Vienna, 2013).
34. Kormányos, A. et al. k-p theory for two-dimensional transition metal dichalcogenide semiconductors. *2D Mater.* **2**, 022001 (2015).
35. Singh, D. J. & Nordström, L. *Planewaves, Pseudopotentials, and the LAPW Method* (Springer, New York, 2006).
36. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
37. Steinhoff, A., Rösner, M., Jahnke, F., Wehling, T. O. & Gies, C. Influence of excited carriers on the optical and electronic properties of MoS₂. *Nano Lett.* **14**, 3743–3748 (2014).
38. Mootz, M., Kira, M. & Koch, S. W. Sequential build-up of quantum-optical correlations. *J. Opt. Soc. Am. B* **29**, A17–A24 (2012).
39. Kira, M. Hyperbolic Bloch equations: atom-cluster kinetics of an interacting Bose gas. *Ann. Phys.* **356**, 185–243 (2015).
40. Mootz, M., Kira, M. & Koch, S. W. Pair-excitation energetics of highly correlated many-body states. *New J. Phys.* **15**, 093040 (2013).
41. Kira, M. & Koch, S. W. Many-body correlations and excitonic effects in semiconductor spectroscopy. *Prog. Quantum Electron.* **30**, 155–296 (2006).
42. Kira, M. Coherent quantum depletion of an interacting atom condensate. *Nat. Commun.* **6**, 6624 (2015).
43. Smith, R. P. et al. Extraction of many-body configurations from nonlinear absorption in semiconductor quantum wells. *Phys. Rev. Lett.* **104**, 247401 (2010).
44. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).



Extended Data Fig. 1 | Sample orientation. **a**, Azimuthal scan of the second-harmonic intensity polarized parallel to the excitation pulse, $I_{\text{SHG},||}$ (blue curve), revealing the armchair direction at a crystal angle of $\varphi = 30^\circ$. The dashed line marks the expected scaling proportional to $\sin^2(3\varphi)$. Around the polar diagram, the hexagonal Brillouin zone of WSe_2 is depicted with the high-symmetry points. **b**, Optical microscope image of the exfoliated monolayer on the visco-elastic gel film used for exfoliation. Areas appearing in lighter grey are few-layer tungsten diselenide. **c**, Monolayer sample after transfer to a diamond substrate. The contrast of this image has been enhanced to improve the visibility of the atomically thin WSe_2 film. The red arrows mark the same edge in **b** and **c**, which has been identified as the zigzag direction using the SHG scan.



Extended Data Fig. 2 | Polarization of subcycle sideband emission.

Circularly polarized 10-fs near-infrared (NIR) pulses (polarization-resolved intensity depicted as black spheres) excite valley-polarized electron-hole pairs in a monolayer of tungsten diselenide. Simultaneously, an atomically strong terahertz wave is applied in the zigzag direction and may transfer electrons and holes to the non-excited K' valley. The high-order sideband emission resulting from coherent electron-hole collisions driven by the most intense half-cycle is measured to have an elliptical polarization (blue spheres), and contains contributions from the opposite valley. Our quantum theory reproduces this polarization state (red curve) and reveals a transfer yield of 66% to the initially unexcited K' valley.

An optical–frequency synthesizer using integrated photonics

Daryl T. Spencer^{1*}, Tara Drake¹, Travis C. Briles^{1,2}, Jordan Stone^{1,2}, Laura C. Sinclair¹, Connor Fredrick^{1,2}, Qing Li³, Daron Westly³, B. Robert Ilic³, Aaron Bluestone⁴, Nicolas Volet⁴, Tin Komljenovic⁴, Lin Chang⁴, Seung Hoon Lee⁵, Dong Yoon Oh⁵, Myoung-Gyun Suh⁵, Ki Youl Yang⁵, Martin H. P. Pfeiffer⁶, Tobias J. Kippenberg⁶, Erik Norberg⁷, Luke Theogarajan⁴, Kerry Vahala⁵, Nathan R. Newbury¹, Kartik Srinivasan³, John E. Bowers⁴, Scott A. Diddams^{1,2*} & Scott B. Papp^{1,2*}

Optical-frequency synthesizers, which generate frequency-stable light from a single microwave-frequency reference, are revolutionizing ultrafast science and metrology, but their size, power requirement and cost need to be reduced if they are to be more widely used. Integrated-photonics microchips can be used in high-coherence applications, such as data transmission¹, highly optimized physical sensors² and harnessing quantum states³, to lower cost and increase efficiency and portability. Here we describe a method for synthesizing the absolute frequency of a lightwave signal, using integrated photonics to create a phase-coherent microwave-to-optical link. We use a heterogeneously integrated III–V/silicon tunable laser, which is guided by nonlinear frequency combs fabricated on separate silicon chips and pumped by off-chip lasers. The laser frequency output of our optical-frequency synthesizer can be programmed by a microwave clock across 4 terahertz near 1,550 nanometres (the telecommunications C-band) with 1 hertz resolution. Our measurements verify that the output of the synthesizer is exceptionally stable across this region (synthesis error of 7.7×10^{-15} or below). Any application of an optical-frequency source could benefit from the high-precision optical synthesis presented here. Leveraging high-volume semiconductor processing built around advanced materials could allow such low-cost, low-power and compact integrated-photonics devices to be widely used.

The electronics revolution that began in the mid-twentieth century was driven in part by advances related to the synthesis of radio and microwave-frequency signals for applications in radar, navigation and communications systems. This formed a foundation for more recent technologies of wide impact, such as the Global Positioning System and cellular communications. Direct-digital synthesis now operates at >10 GHz rates with watt-scale power. Despite the ubiquity of electronic synthesis, no comparable technology existed for electromagnetic signals in the optical domain until the introduction of the self-referenced optical-frequency comb^{4,5}. An optical-frequency comb can provide the critical phase-coherent link between microwave and optical domains, with an output consisting of an array of optical modes having frequencies given exactly by $\nu_n = n f_{\text{rep}} + f_{\text{ceo}}$, where f_{rep} and f_{ceo} are microwave frequencies and n is an integer. Over the past two decades, optical-frequency synthesizers using mode-locked-laser frequency combs have been demonstrated^{6,7}. The optical-synthesizer output, derived from a reference clock, is invaluable for coherent light detection and ranging⁷, atomic and molecular spectroscopy and optical communications. Optical-frequency-comb technology has also matured so that a typical erbium-fibre comb system requires approximately 2 W of optical pump power⁸.

A new opportunity for chip-integrated optical-frequency synthesis has emerged with development in heterogeneously integrated photonics⁹ and photonic-chip microresonator frequency combs, or microcombs^{10–17}. Microresonators pumped by a continuous-wave

(CW) laser generate a parametric four-wave mixing comb in dielectric media. Relying on waveguide confinement and high nonlinearity of the integrated photonics, microresonators provide a route to comb generation with only milliwatts of input power¹⁷ and high pump-conversion efficiency¹⁸. Precise waveguide group-velocity dispersion (GVD) control¹⁹, combined with the realization of low-noise dissipative Kerr solitons (DKSs)^{20–22}, has led to octave-spanning optical spectra with dispersive waves^{23–25} to enhance the signal-to-noise ratio in microcomb carrier-envelope-offset frequency (f_{ceo}) detection^{26–28}. In parallel, through heterogeneous integration it has become possible to seamlessly combine active and passive components, such as semiconductor lasers and amplifiers, electro-optic modulators, passive waveguides, photodiodes and complementary metal–oxide–semiconductor (CMOS) electronics on a silicon-chip platform⁹, and specifically to implement phase-locking of integrated lasers to microcombs^{29,30}. Our work makes use of Kerr-soliton frequency combs and silicon photonics to realize optical-frequency synthesis derived phase-coherently from an electronic clock.

Mirroring the framework of most traditional optical and microwave synthesizers, our system is composed of a tunable laser oscillator that we phase-lock to a stabilized microcomb reference. Figure 1a presents the concept of a future integrated synthesizer, and Fig. 1b indicates the connections between the integrated tunable laser and the chip-based Kerr-comb components that are used in this work. We use the C-band tunability, narrow linewidth and rapid frequency control of a III–V/silicon ring-resonator⁹ laser as the synthesizer output, and the phase-coherent microwave-to-optical connection of a fully stabilized DKS frequency comb. The DKS dual comb consists of an octave-bandwidth, silicon nitride comb with 1 THz mode spacing and a C-band-spanning, fused-silica comb with 22 GHz mode spacing. By phase-stabilizing both comb spacings ($f_{\text{rep,THz}}$ and $f_{\text{rep,GHz}}$) and the silicon nitride comb's offset frequency, $f_{\text{ceo,THz}}$, we establish the precise factor of 19,403,904 phase-coherent multiplication from 10 MHz to the optical domain. With this tunable-laser and frequency-comb system, we demonstrate synthesis across a 4-THz segment of the C-band by programming and dynamically stepping the output frequency; see Fig. 1c–e. As the role of any synthesizer is to output a phase-coherently multiplied version of the input clock, we characterize the optical synthesizer primarily through its fluctuations with an out-of-loop frequency comb derived from the same clock. A fully integrated synthesizer, realized by using, improving and connecting the chip components that we describe, would be a powerful tool for many applications (see Methods).

To demonstrate the optical-frequency synthesizer, we carry out a series of experiments characterizing its output frequency. Standard spectrometer or interferometer measurements readily verify system performance at the megahertz (or 10^{-8}) level. By measuring the synthesizer with an auxiliary self-referenced erbium-fibre comb, we constrain the frequency error between the output and the synthesizer's

¹Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO, USA. ²Department of Physics, University of Colorado, Boulder, CO, USA. ³Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, MD, USA. ⁴University of California Santa Barbara, Santa Barbara, CA, USA. ⁵California Institute of Technology, Pasadena, CA, USA. ⁶Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. ⁷Aurion Inc., Goleta, CA, USA. *e-mail: daryl.spencer@nist.gov; scott.papp@nist.gov

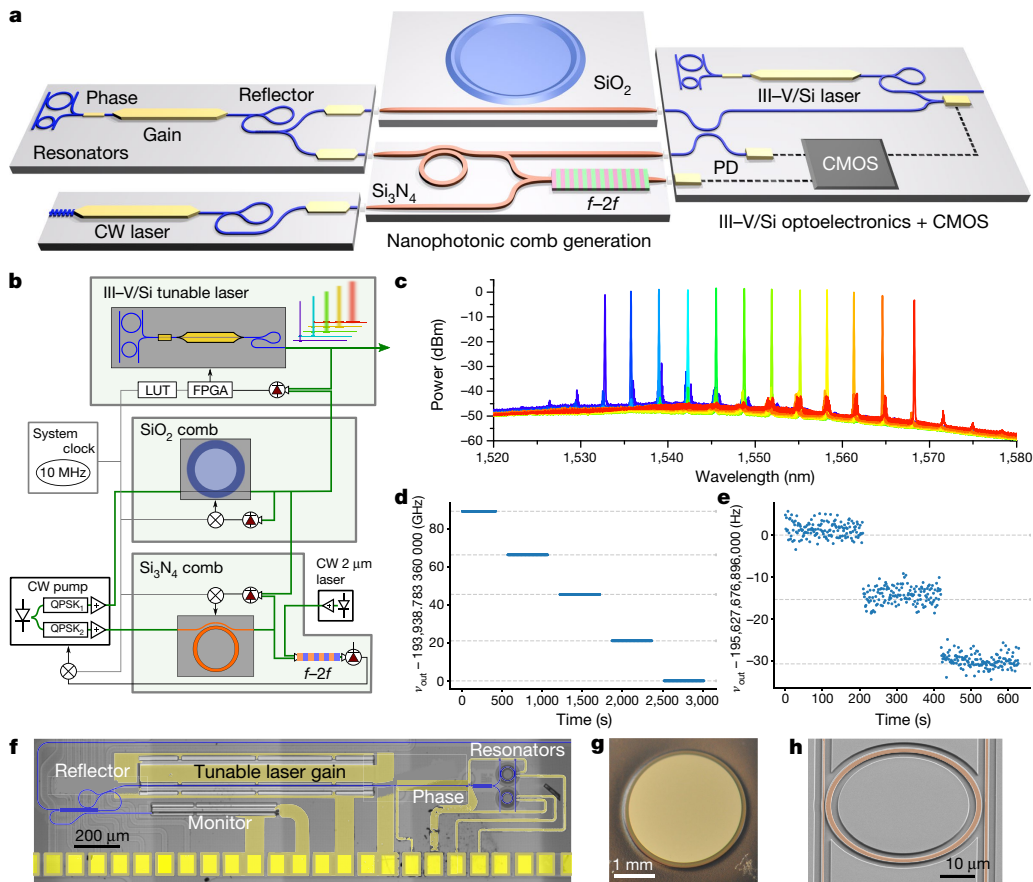


Fig. 1 | Accurate optical synthesis with an integrated laser and DKS dual-comb system. **a**, Conceptual integrated optical-frequency synthesizer with digital control and $f-2f$ stabilization, using the microcombs and tunable laser of this work. PD, photodetector. **b**, Our optical synthesizer is composed of an integrated tunable laser and chip-based Kerr-comb generators. Green boxes indicate the tabletop subsystems including the chips, and how they connect. The CW pumping laser for some experiments is a second integrated laser; see Methods. The tunable laser is synthesized by phase-locking to the stabilized combs, using a look-up table (LUT) and FPGA. QPSK, quadrature phase shift key modulator.

setpoint to <1.5 Hz. Beyond demonstration of the integrated-photonics architecture, the core result of our work is verification that the synthesizer offers sufficient phase control and synchronization in microwave-to-optical conversion (as do the auxiliary comb and our frequency-counting electronics) to reveal a stable phase correlation between the CW output and the radiofrequency (RF) clock. Hence, the statistical fluctuations that lead to the synthesizer's instability, and our measurement of these, offer the complete description of the synthesizer's frequency performance.

The chip-based integrated components of the synthesizer—the tunable laser (Fig. 1f) and DKS frequency combs (Fig. 1g and h)—and their key connections with non-integrated components are emphasized in Fig. 1b. An external cavity pump laser is used to generate both of the DKS combs, using independent control with single-sideband frequency shifters and erbium amplification for each comb. An octave-spanning single-pulse soliton is generated in a Si₃N₄ planar waveguide-coupled resonator. In addition to the anomalous GVD profile, waveguide-dispersion engineering creates dispersive-wave peaks in optical power that appear at 999 nm and 2,190 nm, owing to the zero-integrated GVD starting from the pump wavelength. With a radius of 23 μm, the threshold for octave-spanning spectra is brought to below 50 mW of on-chip pump power²⁴, at the expense of a $f_{\text{rep,THz}}$ of 1.014 THz that cannot be easily photodetected and reduced to a microwave frequency with conventional electronics. Rather, we rely on a second frequency comb to bridge the gap between Si₃N₄ THz comb modes.

c, Optical spectra of the laser across 32 nm. **d, e**, Measurements of the synthesizer output as it is stepped. The data indicate the deviation between the synthesizer output ν_{out} and its setpoint for mode-hopping across the 22-GHz SiO₂ modes (**d**) and for application of precise frequency steps of 15.36 Hz (**e**). **f**, Scanning electron microscope (SEM) image of the heterogeneous III-V/Si tunable laser with false colour electrodes (yellow) and waveguides (blue). **g**, Photograph of the SiO₂-based wedge microresonator. **h**, SEM image of the Si₃N₄ THz resonator with false colour imposed on the waveguide regions.

To do this, an SiO₂ wedge-based whispering-gallery-mode resonator with a quality factor (Q) of 180 million is used to create a DKS frequency comb at $f_{\text{rep,GHz}} \approx 22$ GHz (ref. ²²). As the threshold for soliton-comb generation scales inversely to both the repetition rate and Q^2 , use of an SiO₂ device is important for low-power operation. The repetition frequency of 22 GHz is photodetected and phase-locked to the RF clock. This first step in the microwave-to-optical frequency chain (Fig. 2a) from $f_{\text{clk}} = 10$ MHz to 22 GHz partially stabilizes the SiO₂ reference comb to guide tunable laser synthesis; see Fig. 2b. The second step is detection of the 1.014 THz frequency spacing between Si₃N₄ comb teeth, which we accomplish using the 46th relative comb line from the SiO₂ comb. Operationally, we measure $f_{\text{rep,THz}}$ by detecting the optical heterodyne beat note between the two combs 1 THz away from the pump. We phase-lock this signal to a synthesized radiofrequency, $f_1 = \alpha f_{\text{clk}}$ (where α is the ratio of two integers), after removing the relative contributions from the single-sideband frequency shifters and feeding back to the frequency of the Si₃N₄ pump laser^{31,32}. Thus, we stabilize $f_{\text{rep,THz}}$ and transfer the f_{clk} stability to 1.014 THz. The frequency of each of the Si₃N₄ THz comb lines with negative offset frequency and mode number $N = 192$ is then given by:

$$\begin{aligned} \nu_{\text{THz}} &= N f_{\text{rep,THz}} - f_{\text{ceo,THz}} \\ \nu_{\text{THz,pump}} &= 192(46 f_{\text{rep,GHz}} + \alpha f_{\text{clk}}) - f_{\text{ceo,THz}} \end{aligned} \quad (1)$$

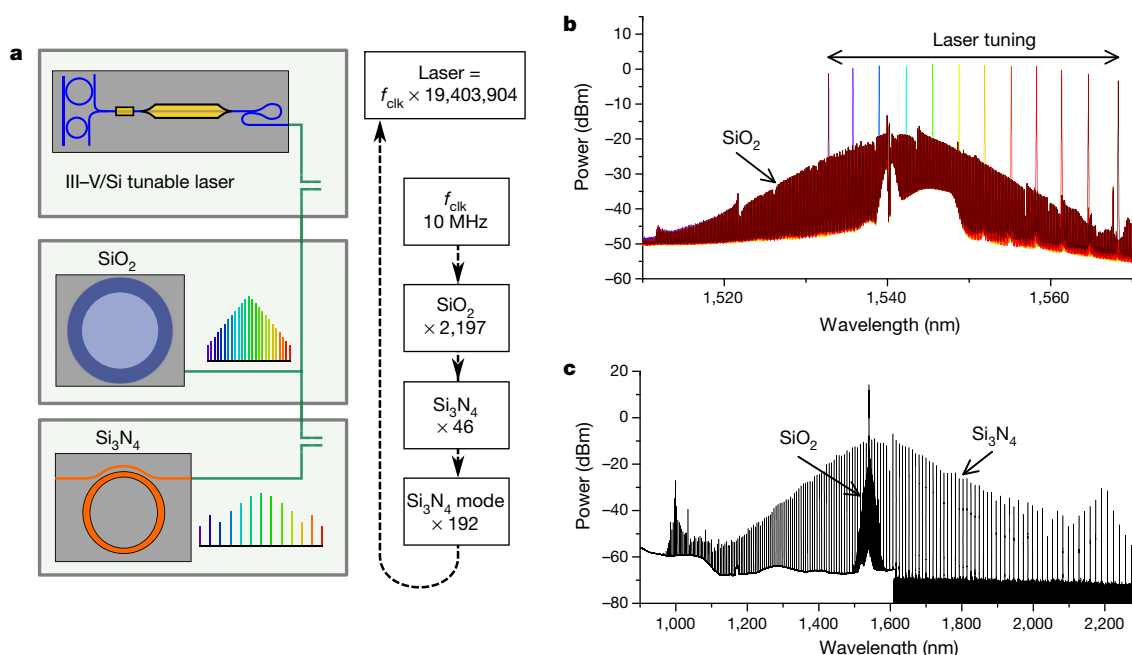


Fig. 2 | Optical spectra of the integrated devices. a, Schematic diagram of spectral combination with the integrated devices, and the frequency chain used to multiply the 10-MHz clock to the optical domain. **b**, Combined spectrum of the SiO_2 22-GHz wedge microcomb and the heterogeneously

integrated III-V/Si tunable laser in the telecommunication C-band. **c**, Combined spectrum of the octave-spanning Si_3N_4 THz microcomb and the 22-GHz SiO_2 wedge microcomb, as measured on two optical spectrum analysers.

Next, $f_{\text{ceo,THz}}$ locking is achieved by using the octave-spanning relationship of the THz lines at 1,998 nm and dispersive wave peak at 999 nm (Fig. 2c). To aid $f-2f$ self-referencing (which enables determination of the absolute frequency of each comb line), an independent diode laser and thulium-doped fibre amplifier at 1,998 nm supply 9 mW to a waveguide periodically poled lithium niobate (PPLN) device to

achieve 34 dB signal-to-noise ratio (SNR) on $f_{\text{ceo,THz}}$. Similar monolithic second-harmonic generation and amplifier technologies have been demonstrated and could be integrated with our system (see Methods). After detecting two heterodyne beats with the THz comb, f_{999} and f_{1998} , each beat note is digitally divided by 64 and 32, respectively, and frequency mixing yields an $f_{\text{ceo,THz}}$ signal, $f_{\text{ceo,THz}}/64 = f_{999}/64 - f_{1998}/32$.

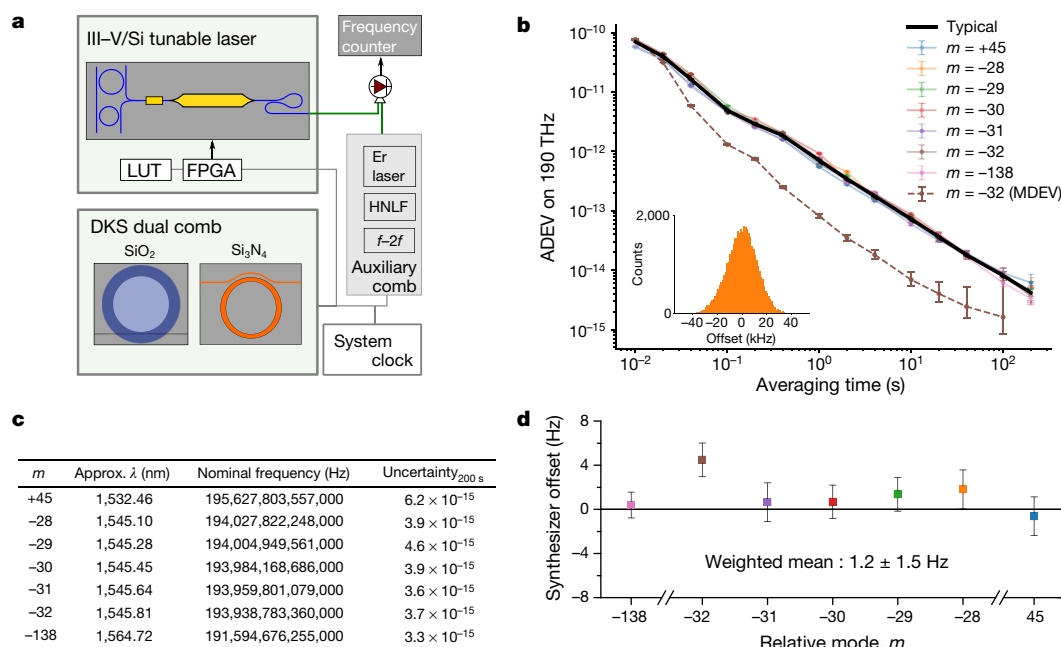


Fig. 3 | Stable optical synthesis with out-of-loop verification. a, Tunable laser locking, and frequency counting with the auxiliary comb. HNLFF, highly nonlinear fibre. **b**, Measured overlapping Allan deviation (ADEV) and modified Allan deviation (MDEV) of the frequency synthesizer. In comparing 10-ms counter-gate time acquisitions, the $1/\tau$ slope is consistent with a stable, phase-locked synthesizer, and the histograms of 500 s of data (inset for relative mode $m = -28$ only) show a Gaussian

profile. Error bars indicating 95% confidence intervals are derived using flicker noise estimates (see Methods). **c**, Table of nominal frequencies and uncertainty at 200 s as the synthesizer is stepped across the C-band. **d**, Overview of the accuracy and precision of the synthesizer frequency. The ADEV at 100 s is used to estimate the uncertainty of each synthesizer output, and the weighted mean of the seven data points is reported with a 95% (t distribution) confidence interval.

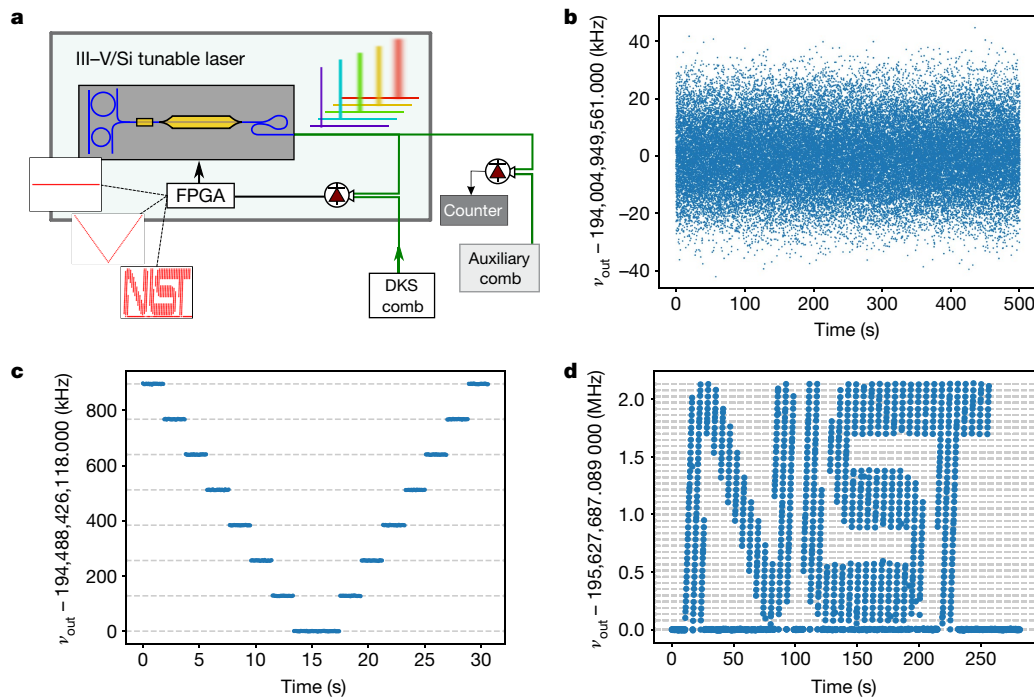


Fig. 4 | Arbitrary control of the optical-frequency synthesizer. a, Step-wise control of the tunable-laser offset phase-lock to the DKS comb and frequency counting. **b**, Deviation between the synthesizer output ν_{out} and constant setpoint for 500 s at a 10-ms gate time. **c**, Bidirectional linear ramp of the synthesizer via step control of the laser offset PLL setpoint

Phase-locking this signal to a radiofrequency $f_2 = \beta f_{\text{clk}}$, through feedback to the Si_3N_4 pump power, completes the transfer of stability from f_{clk} to all the THz comb lines spanning 130 THz to 300 THz.

The dual-stabilized combs serve as the backbone to guide the heterogeneously integrated III–V/Si tunable laser for arbitrary optical-frequency synthesis across the C-band. The tunable laser consists of InGaAsP multiple-quantum-well epitaxial material that is wafer-bonded onto a lithographically patterned silicon-on-insulator wafer⁹. Bias heaters integrated on the laser’s Si-based resonant reflectors and phase section are used to shift the lasing wavelength for initial alignment to the comb lines. By using Si waveguides that have low loss relative to standard telecommunication-grade InP waveguide technology, reduced linewidths of about 300 kHz are achieved. The combined optical spectrum of the SiO_2 comb and integrated laser’s tuning range is shown in Fig. 2b. Heterodyning with the DKS dual-comb signal at a relative mode m from the pump creates a signal, $f_{\text{beat}}^{\text{laser}}$, for input to a field-programmable gate array (FPGA)-based phase-locked-loop (PLL, Fig. 3a) with a local oscillator of $f_3 = \gamma f_{\text{clk}}$, and digital division of 512. This final laser lock to the DKS dual comb produces a fully stabilized, tunable synthesizer output,

$$\begin{aligned} \nu_{\text{out}} &= \nu_{\text{THz,pump}} + m f_{\text{rep,GHz}} + f_{\text{beat}}^{\text{laser}} \\ &= f_{\text{clk}} [192(46 \times 2,197 + \alpha) + 2,197m - 64\beta + 512\gamma] \end{aligned} \quad (2)$$

This expression shows that the output of our integrated-photonics synthesizer is uniquely and precisely defined relative to the input clock frequency in terms of user-chosen integers and ratios of integers (α, β, γ).

Agile tuning across SiO_2 comb lines (varying m) and hertz-level tuning resolution on the same comb line (varying γ) have already been presented in Fig. 2b and Fig. 1c, demonstrating synthesizer operation. To explore our synthesizer’s phase coherence, we perform an out-of-loop optical-frequency characterization by heterodyning ν_{out} against an auxiliary erbium-fibre laser frequency comb that is fully stabilized to the same f_{clk} . Figure 3 shows results from a study of the tunability and

(100-ms gate). **d**, Arbitrary frequency control of the synthesizer across 40 frequency setpoints to write “NIST”. A 30-ms gate time is used to oversample each frequency setpoint by 5 (150-ms pause per point), and every fifth data point is displayed.

phase-locked operation of the synthesizer across all comb frequencies by locking to five adjacent SiO_2 comb lines, and to the highest and lowest wavelengths of the laser tuning range. Overlapping Allan deviation (ADEV) analysis of the counted beat notes against the auxiliary comb show the instability improving as $<10^{-12}/\tau$ for all recorded averaging times τ , and reaching an average instability of $(4.2 \pm 0.4) \times 10^{-15}$ at 200 s (Fig. 3b, c). More sophisticated triangular averaging analysis using the modified Allan deviation (MDEV) yields an order of magnitude better instability of $(9.2 \pm 1.4) \times 10^{-14}$ at 1 s. Still, the $1/\tau$ dependence of the ADEV data, which characterizes the fluctuations of the optical-frequency synthesizer, indicates the stable phase relationship between the RF clock and the synthesized optical frequencies. Moreover, the synthesizer performance is consistent with the hydrogen-maser RF clock used in the experiments, indicating that our phase locks of the tunable laser, the Kerr combs and the auxiliary erbium-fibre comb contribute negligible noise. This is the most fundamental metric of an optical synthesizer. From the mean values of the measured beats with the auxiliary comb, we can further analyse potential deviations of the synthesizer output from equation (2). Data compiled from the seven experiments are shown in Fig. 3d with 100-s ADEV error bars plotted, and the weighted mean of all data sets with a 95% confidence interval is (1.2 ± 1.5) Hz. Thus, based on these initial data, we conclude that our integrated-photonics optical synthesizer accurately reproduces the input clock reference within an uncertainty of 7.7×10^{-15} , competitive with commercial optical synthesizers (5×10^{-13} instability at 1 s and accurate to 10^{-14} at 120 s).

To demonstrate the tunability of the optical-frequency synthesizer, we perform two different types of tuning while the laser is locked to the stabilized comb system (Fig. 4). As a baseline, without changing the setpoint of the tunable laser phase-lock, the raw data of the counted auxiliary comb beat note are shown in Fig. 4b after subtraction of the nominally expected frequency for 500 s. We then apply a bidirectional linear ramp over eight levels with a 2-s pause at each level to ensure successful locking (Fig. 4c). Finally, we programme a series of setpoint frequencies to the FPGA PLL box to write out the National Institute of Standards and Technology (NIST) logo (Fig. 4d). Excellent agreement

is found between the expected offset frequencies and the counted beat-note frequencies for all cases, illustrating good dynamic control of the synthesizer.

In summary, the experiments that we present, performed with an optical-frequency synthesizer constructed from integrated photonics, demonstrate that this technology has achieved the high precision and accuracy that formerly has been confined to tabletop mode-locked laser frequency-comb devices. For further integration of the laser and Kerr combs used in our experiments, targeted improvements should be made to increase microresonator Q for lower-power operation, to improve the intensity of the Si_3N_4 comb dispersive waves for $f-2f$ stabilization, and to improve the efficiency of second-harmonic generation, guided by the applications that are envisaged for the device.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0065-7>.

Received: 21 August 2017; Accepted: 22 January 2018;

Published online 25 April 2018.

- Rumley, S. et al. Silicon photonics for exascale systems. *J. Lightwave Technol.* **33**, 547–562 (2015).
- Purdy, T. P., Grutter, K. E., Srinivasan, K. & Taylor, J. M. Quantum correlations from a room-temperature optomechanical cavity. *Science* **356**, 1265–1268 (2017).
- O'Brien, J. L., Furusawa, A. & Vučković, J. Photonic quantum technologies. *Nat. Photonics* **3**, 687–695 (2009).
- Hall, J. L. Nobel Lecture: Defining and measuring optical frequencies. *Rev. Mod. Phys.* **78**, 1279–1295 (2006).
- Hänsch, T. W. Nobel Lecture: Passion for precision. *Rev. Mod. Phys.* **78**, 1297–1309 (2006).
- Jost, J. D., Hall, J. L. & Ye, J. Continuously tunable, precise, single frequency optical signal generator. *Opt. Express* **10**, 515–520 (2002).
- Giorgetta, F. R., Coddington, I., Baumann, E., Swann, W. C. & Newbury, N. R. Fast high-resolution spectroscopy of dynamic continuous-wave laser sources. *Nat. Photonics* **4**, 853–857 (2010).
- Sinclair, L. C. et al. Operation of an optically coherent frequency comb outside the metrology lab. *Opt. Express* **22**, 6996–7006 (2014).
- Komljenovic, T. et al. Heterogeneous silicon photonic integrated circuits. *J. Lightwave Technol.* **34**, 20–35 (2016).
- Del'Haye, P. et al. Optical frequency comb generation from a monolithic microresonator. *Nature* **450**, 1214–1217 (2007).
- Savchenkov, A. A. et al. Tunable optical frequency comb with a crystalline whispering gallery mode resonator. *Phys. Rev. Lett.* **101**, 093902 (2008).
- Kippenberg, T. J., Holzwarth, R. & Diddams, S. A. Microresonator-based optical frequency combs. *Science* **332**, 555–559 (2011).
- Moss, D. J., Morandotti, R., Gaeta, A. L. & Lipson, M. New CMOS-compatible platforms based on silicon nitride and Hydex for nonlinear optics. *Nat. Photonics* **7**, 597–607 (2013).
- Grudinin, I. S., Yu, N. & Maleki, L. Generation of optical frequency combs with a CaF_2 resonator. *Opt. Lett.* **34**, 878–880 (2009).
- Ferdous, F. et al. Spectral line-by-line pulse shaping of on-chip microresonator frequency combs. *Nat. Photonics* **5**, 770–776 (2011).
- Papp, S. B. & Diddams, S. A. Spectral and temporal characterization of a fused-quartz-microresonator optical frequency comb. *Phys. Rev. A* **84**, 053833 (2011).
- Li, J., Lee, H., Chen, T. & Vahala, K. J. Low-pump-power, low-phase-noise, and microwave to millimeter-wave repetition rate operation in microcombs. *Phys. Rev. Lett.* **109**, 233901 (2012).
- Cole, D. C., Lamb, E. S., Del'Haye, P., Diddams, S. A. & Papp, S. B. Soliton crystals in Kerr resonators. *Nat. Photonics* **11**, 671–676 (2017).
- Okawachi, Y. et al. Bandwidth shaping of microresonator-based frequency combs via dispersion engineering. *Opt. Lett.* **39**, 3535–3538 (2014).
- Herr, T. et al. Temporal solitons in optical microresonators. *Nat. Photonics* **8**, 145–152 (2014).
- Leo, F. et al. Temporal cavity solitons in one-dimensional Kerr media as bits in an all-optical buffer. *Nat. Photonics* **4**, 471–476 (2010).
- Yi, X., Yang, Q.-F., Yang, K. Y., Suh, M.-G. & Vahala, K. Soliton frequency comb at microwave rates in a high- Q silica microresonator. *Optica* **2**, 1078–1085 (2015).
- Brasch, V. et al. Photonic chip-based optical frequency comb using soliton Cherenkov radiation. *Science* **351**, 357–360 (2016).
- Li, Q. et al. Stably accessing octave-spanning microresonator frequency combs in the soliton regime. *Optica* **4**, 193–203 (2017).
- Pfeiffer, M. H. P. et al. Octave-spanning dissipative Kerr soliton frequency combs in Si_3N_4 microresonators. *Optica* **4**, 684–691 (2017).
- Jost, J. D. et al. Counting the cycles of light using a self-referenced optical microresonator. *Optica* **2**, 706–711 (2015).
- Del'Haye, P. et al. Phase-coherent microwave-to-optical link with a self-referenced microcomb. *Nat. Photonics* **10**, 516–520 (2016).
- Brasch, V., Lucas, E., Jost, J. D., Geiselmann, M. & Kippenberg, T. J. Self-referenced photonic chip soliton Kerr frequency comb. *Light Sci. Appl.* **6**, e16202 (2017).
- Arafin, S. et al. Power-efficient Kerr frequency comb based tunable optical source. *IEEE Photonics J.* **9**, 6600814 (2017).
- Arafin, S. et al. Towards chip-scale optical frequency synthesis based on optical heterodyne phase-locked loop. *Opt. Express* **25**, 681–695 (2017).
- Del'Haye, P., Arcizet, O., Schliesser, A., Holzwarth, R. & Kippenberg, T. J. Full stabilization of a microresonator-based optical frequency comb. *Phys. Rev. Lett.* **101**, 053903–053904 (2008).
- Papp, S. B. et al. Microresonator frequency comb optical clock. *Optica* **1**, 10–14 (2014).

Acknowledgements We thank Srico, Inc. for use of the waveguide PPLN device, Aurion Inc. for use of the III–V/Si tunable laser, and D. Hickstein, T. Dunker, A. Wallin, D. Carlson and Z. Newman for comments on the experiment. N.V. acknowledges support from the Swiss National Science Foundation (SNSF). This research is supported by the Defense Advanced Research Projects Agency DODOS program and NIST. We thank R. Lutwak and the DODOS program management team for discussions throughout the experiment.

Reviewer Information Nature thanks M. Lipson, D. Moss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author Contributions D.T.S., T.D., T.C.B. and J.S. contributed equally to performing the system measurements and analysing the experimental results. D.T.S., S.A.D. and S.B.P. prepared the manuscript. The integrated devices were fabricated and tested by Q.L., D.W., B.R.I. and K.S. (Si_3N_4); A.B., N.V., T.K., L.C. and E. N. (III–V/Si); and S.H.L., D.Y.O., M.S., K.Y.Y. and K.V. (SiO_2). N.V., L.C.S., C.F., M.H.P.F. and A.B. provided measurement support. T.J.K., E.N., K.V., K.S., N.R.N., L.T., J.E.B., S.A.D. and S.B.P. supervised and led the scientific collaboration. This work is an official contribution of the NIST; not subject to copyright in the United States. The use of trade names is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Competing interests The authors declare no competing financial interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0065-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to D.T.S. or S.B.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Device and experimental details. The heterogeneously integrated III–V/Si device includes a tunable laser and a semiconductor optical amplifier (SOA). At room temperature, the laser emits up to 4 mW CW power, and the SOA provides an on-chip small-signal gain >10 dB. The laser contains a gain section, a phase section and two microresonators designed for high quality factor. The gain section and the SOA consist of electrically pumped InP-based quantum wells heterogeneously integrated on a Si waveguide⁹. Thermal heating of the Si microresonators and passive phase section is performed with current injection to metal heaters above the waveguides. By intentionally mismatching the radii of the microresonators to make use of the Vernier effect, we can use a narrowband intracavity optical filter to select the desired longitudinal mode for lasing with high side-mode suppression ratio³³. Precise wavelength tuning and linewidth narrowing is performed by heating the phase section (Extended Data Fig. 1). Phase-locking of the laser to the microcomb is performed by electronically dividing the beat note by 512 and using FPGA-based digital PLL + PI²D feedback (that is, a proportional–integral–derivative controller with two-stage integration) to the gain section of the laser⁸. Other works have also demonstrated high-bandwidth optical-PLL phase locks to frequency combs^{29,30}. In the current system, the tunable laser can lock to either the SiO₂ or Si₃N₄ comb lines, which we have shown for $m = -138$, or 3 THz red of the pump laser. The DC linear tuning coefficient of the tunable laser is approximately 200 MHz mA⁻¹.

A commercial external cavity diode laser is used as the shared pump for both microresonator comb generators in all the synthesizer experiments. The output of a 3-dB splitter goes to separate LiNbO₃ single-sideband modulators and erbium-doped fibre amplifiers for each device. Frequency detuning from each microcomb resonance for soliton generation is controlled with an amplified voltage-controlled oscillator and arbitrary waveform generator that produces a voltage ramp. Although complex soliton crystals¹⁸ can form in these devices, single solitons are generated through linear voltage ramps of 5 GHz in 100 ns and 100 MHz in 3 μ s, for the Si₃N₄ and SiO₂ microcombs, respectively³⁴. Once initiated, feedback to each voltage-controlled oscillator controls $f_{\text{rep,GHz}}$ and $f_{\text{rep,THz}}$ for the appropriate device. Intensity modulation on the Si₃N₄ microcomb to control $f_{\text{ceo,THz}}$ is performed with a free-space acousto-optic modulator, although on-chip SOAs are expected to be viable as well. Lensed fibres with a 2.5- μ m spot size are used to couple light on and off the Si₃N₄ chip with 7 dB of insertion loss per facet. During operation, the on-chip pump power for the Si₃N₄ microcomb is about 160 mW, or ten times the threshold for soliton generation. Tapered single-mode fibre is used to couple 80 mW to the SiO₂ microcomb for soliton generation at 12 times the soliton threshold. Recent results show that this platform can be integrated with Si₃N₄ bus waveguides³⁵. An offset Pound–Drever–Hall lock is required after ramping to keep the SiO₂ pump frequency at 22 MHz red detuned from resonance³⁶.

During operation of the optical-frequency synthesizer, the separate single-sideband modulators for each microcomb device create a detectable offset in pump frequencies, about 5 GHz in our experiment. This is readily subtracted from or added to the necessary heterodyne beat notes in the system using an electronic frequency mixer, specifically after $f_{\text{rep,THz}}$ detection between comb lines and after the III–V/Si laser heterodyne with the DKS comb. The calibrated gain sign of the tunable-laser feedback loop ensures that the tunable laser is on the appropriate side of the SiO₂ comb modes when electronically subtracting or adding this offset, and knowledge of the absolute difference in pump frequencies is not required for accurate optical-frequency synthesis. We observe non-zero synthesis error when the SNR of any heterodyne beat falls well below the optimal level of 30 dB, but measurements reported here were acquired with sufficient SNR. We also observe and minimize contributions from out-of-loop optical and electrical path lengths, alignment drift, and glitches during long acquisitions. The RF synthesis and phase-locking electronics used in the experiments are benchtop scale, but in the future would make use of CMOS integration³⁷.

Auxiliary comb details and frequency counting. The auxiliary comb used for out-of-loop verification of the optical-frequency synthesizer consists of a 250-MHz erbium-fibre mode-locked laser frequency comb³⁸. The laser output is amplified and spectrally broadened to an octave to enable self-referenced detection of the

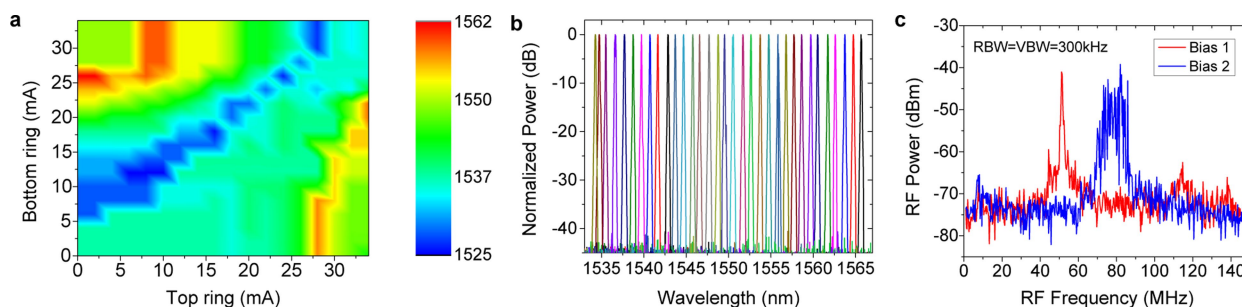
carrier envelope offset frequency, f_{ceo} . The fourth harmonic of f_{rep} is phase-locked to a reference synthesizer at 999.999 544 MHz, and f_{ceo} is electronically divided by 8 and phase-locked to another synthesizer at 20 MHz. Both of these synthesizers are referenced to the same f_{clk} that is the input to the integrated-photonics synthesizer, yielding a comb against which any frequency of the microcomb or tunable laser output can be compared.

The beat-note frequency between the integrated-photonics synthesizer and the erbium-fibre frequency comb is amplified and bandpass-filtered (45 MHz bandwidth), after which a zero-dead-time frequency counter registers the frequency for a fixed gate time. The rectangular binning, or Π -mode, is used during measurement and for the ADEV analysis. The MDEV analysis applies a triangular averaging window to the frequency data for further information on the noise type. With this analysis, a $\tau^{-3/2}$ slope shows the desired white phase noise performance, and deviation from this slope reveals unwanted flicker phase noise contributes to system performance at longer averaging times. Because the degrees of freedom depend on noise type, we take the conservative estimate of flicker phase noise to derive 95% confidence intervals³⁹. The tunable laser PLL also contains an in-loop frequency counter, which showed tight phase-locking of the laser to the microcomb at $<10^{-13}/\tau$, limited by the resolution of the counter. All RF synthesizers in the experimental set-up, auxiliary comb and frequency counter are tied to the same hydrogen maser signal, serving as f_{clk} .

Perspectives and future work. A critical element to operation of the optical synthesizer is the pump laser of the DKS microcombs. We show that the same III–V/Si tunable laser from this work can be used to generate low-noise solitons in the Si₃N₄ microresonator; see Extended Data Fig. 2. Further development is required to stabilize solitons in the high-Q SiO₂ microresonator with the III–V/Si tunable laser, although we observe modulation instability (non-soliton) Kerr combs and their transient decay through the Kerr soliton stability regime. Further technical improvement of the III–V/Si tunable laser would probably permit soliton stabilization. At present, we require an optical power of 80 mW for the SiO₂ comb, 160 mW on-chip for the Si₃N₄ comb and 9 mW on-chip for the PPLN device. In each case, we anticipate improving the chip-device performance to be compatible with available integrated-laser power levels to support further integration of our frequency synthesizer. In future implementations of our optical-frequency synthesizer, technical improvements such as improved on and off chip coupling, long wavelength SOAs⁴⁰ and higher efficiency second-harmonic generation⁴¹ would make the 1,998-nm diode laser unnecessary.

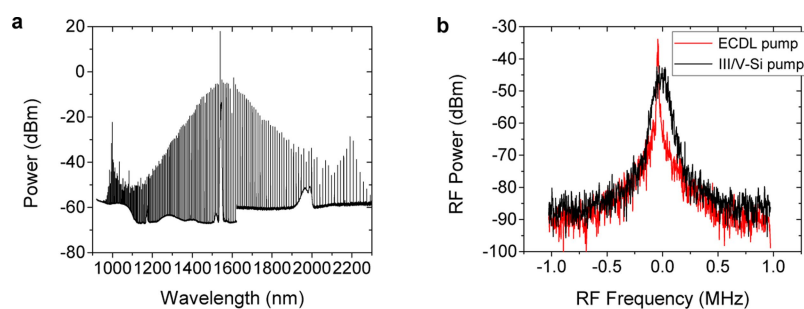
Data availability. The data sets generated and/or analysed during the current study are available from the corresponding authors on reasonable request.

33. Komljenovic, T. et al. Widely tunable narrow-linewidth monolithically integrated external-cavity semiconductor lasers. *IEEE J. Sel. Top. Quantum Electron.* **21**, 214–222 (2015).
34. Briles, T. C. et al. Kerr-microresonator solitons for accurate carrier-envelope-frequency stabilization. Preprint at <https://arxiv.org/abs/1711.06251> (2017).
35. Yang, K. Y. et al. Bridging ultrahigh-Q devices and photonic circuits. *Nat. Photonics* (2018).
36. Stone, J. et al. Thermal and nonlinear dissipative-soliton dynamics in Kerr microresonator frequency combs. Preprint at <https://arxiv.org/abs/1708.08405> (2017).
37. Bluestone, A. et al. Heterodyne-based hybrid controller for wide dynamic range optoelectronic frequency synthesis. *Opt. Express* **25**, 29086–29097 (2017).
38. Ycas, G., Osterman, S. & Diddams, S. A. Generation of a 660–2100 nm laser frequency comb based on an erbium fiber laser. *Opt. Lett.* **37**, 2199–2201 (2012).
39. Greenhall, C. A. & Riley, W. J. in *Proc. PTI 2003*, 267–280 (2003).
40. Volet, N. et al. Semiconductor optical amplifiers at 2.0- μ m wavelength on silicon. *Laser Photonics Rev.* **11**, 1600165 (2017).
41. Chang, L. et al. Thin film wavelength converters for photonic integrated circuits. *Optica* **3**, 531–535 (2016).
42. Srinivasan, S. et al. Coupled-ring-resonator-mirror-based heterogeneous III–V silicon tunable laser. *IEEE Photonics J.* **7**, 2700908 (2015).
43. Del'Haye, P., Papp, S. B. & Diddams, S. A. Hybrid electro-optically modulated microcombs. *Phys. Rev. Lett.* **109**, 263901 (2012).



Extended Data Fig. 1 | Tuning details for III-V/Si laser. **a**, Typical tuning map of the III-V/Si tunable laser's peak wavelength in nanometres versus current applied to each heater above the ring resonators. **b**, Normalized optical spectra showing >40 dB of side-mode suppression ratio across the tuning range. **c**, Typical unlocked RF beat notes between the tunable laser and the auxiliary comb for two different biases of the phase section.

Careful control of the heater is required to reach all wavelengths in the tuning range, and reduction of the laser linewidth (blue to red) through longitudinal mode alignment and the optical feedback effect⁴² is required to achieve the best phase-locking performance to the microcombs. RBW, resolution bandwidth, VBW, video bandwidth.



Extended Data Fig. 2 | Demonstration of pumping the Si_3N_4 THz microcomb with the III-V/Si laser. **a**, Output optical spectrum of the THz microcomb showing dual-dispersive waves, as measured on two optical spectrum analysers. **b**, Comparison of electro-optic repetition rate

detection⁴³ when using the same III-V/Si laser (black) and external cavity diode laser (ECDL, red) from the main experiment to pump the THz microcomb.

Hyperexpandable, self-healing macromolecular crystals with integrated polymer networks

Ling Zhang¹, Jake B. Bailey¹, Rohit H. Subramanian¹ & F. Akif Tezcan^{1,2*}

The formation of condensed matter typically involves a trade-off between structural order and flexibility. As the extent and directionality of interactions between atomic or molecular components increase, materials generally become more ordered but less compliant, and vice versa. Nevertheless, high levels of structural order and flexibility are not necessarily mutually exclusive; there are many biological (such as microtubules^{1,2}, flagella³, viruses^{4,5}) and synthetic assemblies (for example, dynamic molecular crystals^{6–9} and frameworks^{10–13}) that can undergo considerable structural transformations without losing their crystalline order and that have remarkable mechanical properties^{8,14,15} that are useful in diverse applications, such as selective sorption¹⁶, separation¹⁷, sensing¹⁸ and mechanoactuation¹⁹. However, the extent of structural changes and the elasticity of such flexible crystals are constrained by the necessity to maintain a continuous network of bonding interactions between the constituents of the lattice. Consequently, even the most dynamic porous materials tend to be brittle and isolated as microcrystalline powders¹⁴, whereas flexible organic or inorganic molecular crystals cannot expand without fracturing. Owing to their rigidity, crystalline materials rarely display self-healing behaviour²⁰. Here we report that macromolecular ferritin crystals with integrated hydrogel polymers can isotropically expand to 180 per cent of their original dimensions and more than 500 per cent of their original volume while retaining periodic order and faceted Wulff morphologies. Even after the separation of neighbouring ferritin molecules by 50 Å upon lattice expansion, specific molecular contacts between them can be reformed upon lattice contraction, resulting in the recovery of atomic-level periodicity and the highest-resolution ferritin structure reported so far. Dynamic bonding interactions between the hydrogel network and the ferritin molecules endow the crystals with the ability to resist fragmentation and self-heal efficiently, whereas the chemical tailorability of the ferritin molecules enables the creation of chemically and mechanically differentiated domains within single crystals.

Hydrogel polymers present a stark contrast to molecular crystals in that they lack structural order, but are highly elastic and adaptive, can expand considerably and self-heal when equipped with dynamic bonding functionalities^{21,22}. Previously, the isotropic swelling–contraction behaviour of hydrogels has been used to modulate the lattice spacing of colloidal nanoparticle arrays²³, and recently, to expand biological tissue samples and thus facilitate high-resolution fluorescence imaging²⁴. In this study, we examine whether the mechanical properties of hydrogels could be endowed upon molecular crystals. That is, can crystal lattices that are formed by discrete molecules that are connected via specific bonding interactions be mechanically modulated through the integration of polymeric hydrogels? To create hydrogel-expandable molecular crystals, we surmised that the following design parameter conditions should be met: (1) lattices should be mesoporous to enable the hydrogel network to penetrate efficiently and uniformly into the crystals; (2) intermolecular interactions between the constituents of the lattices should be reversible and chemically specific (that is, contain

directional and dynamic bonds), such that they disengage with ease during expansion and re-engage with high fidelity upon contraction; (3) interactions between the constituents of the lattice and the hydrogel network should be extensive to maintain the integrity of the crystal–polymer hybrid at all times and sufficiently dynamic to minimize the build-up of local strain and to enable self-healing.

With these parameters in mind, we arrived at hybrid materials composed of ferritin crystals integrated with the superabsorbent poly(acrylate–acrylamide), or p(Ac–Am), copolymer hydrogels, whose swelling–contraction behaviour can be modulated by the ionic strength and pH²⁵. Ferritin is a 24-meric, quasi-spherical protein with 432 symmetry, an outer diameter of 12 nm, an inner diameter of 8 nm, and a molecular weight²⁶ of more than 500,000 Da. Human heavy-chain ferritin forms highly ordered, face-centred cubic (fcc) crystals that routinely grow to more than 200 µm in size and diffract to less than 2.0 Å. The fcc lattice (Fig. 1a) is characterized by a mesoporous network consisting of cube-shaped, 6-nm-wide chambers (Fig. 1b) that are interconnected by smaller, octahedron-shaped cavities that taper to a pore size of about 2 nm at their narrowest (Fig. 1c), thus fulfilling condition (1). The lattice is formed through highly specific, metal-mediated contacts between neighbouring ferritin molecules (Fig. 1d), which are promoted through the K86Q surface mutation to enable metal coordination²⁷. The absence of any other interprotein contacts means that the entire lattice bonding framework of ferritin molecules can be formed or broken via binding or removal of metal ions (such as Ca²⁺), satisfying condition (2). Finally, ferritin bears a small negative charge, with a zeta potential ranging from –5.5 mV at pH 6.0 to –7.3 mV at pH 7.5 (Extended Data Fig. 1a, b). The exterior surface of ferritin presents a diffuse distribution of both negatively and positively charged residues (Extended Data Fig. 1c), which should enable uniform association with the p(Ac–Am) network through a combination of ionic and H-bonding interactions, thus fulfilling condition (3) (Fig. 1e).

We first examined the efficiency of molecular diffusion and polymerization within ferritin crystals. Diffusion into single ferritin crystals was assessed using the fluorescent tracer rhodamine B by confocal fluorescence microscopy experiments. These experiments showed that a typical crystal (edge length, $l_{\text{edge}} = 50\text{--}250\text{ }\mu\text{m}$) was completely infiltrated by rhodamine B (Extended Data Fig. 2a), which is considerably larger (479 g mol^{-1}) than the Ac and Am molecules (both 71 g mol^{-1}), within 15 min. In a typical preparation of crystal–hydrogel hybrids, ferritin crystals were incubated with polymer precursors (8.625% (w/v) sodium acrylate, 2.5% acrylamide and 0.2% N,N'-methylenebis(acrylamide)) for at least 10 h to ensure their uniform distribution in the lattice interstices. This treatment caused no apparent damage to the crystals (see Supplementary Information for quantification of polymer precursor concentrations inside the crystals). Crystals were then transferred into a solution containing 1% (w/v) ammonium persulfate (APS) and 1% (v/v) tetramethylethylenediamine (TEMED) to initiate free-radical polymerization within the lattice, as well as 4 M sodium chloride (NaCl) to limit swelling during polymerization (Fig. 1e). To assess the kinetics of polymerization inside the crystals, we added 0.3% (w/v)

¹Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA. ²Materials Science and Engineering, University of California, San Diego, La Jolla, CA, USA. *e-mail: tezcan@ucsd.edu

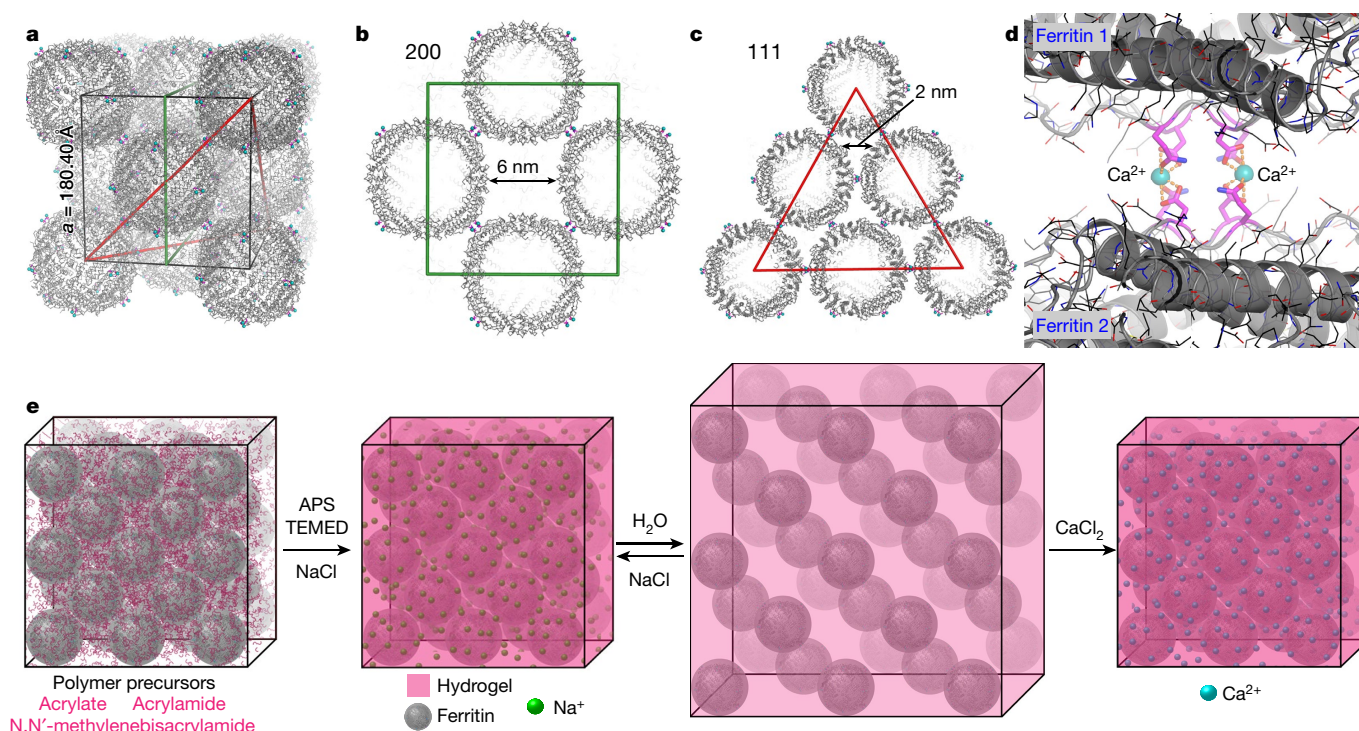


Fig. 1 | Packing arrangement in ferritin crystals and their expansion-contraction mediated by the infused hydrogel network. a–c, The fcc packing arrangement of ferritin crystals (Protein Data Bank identifier, PDB ID, 6B8F). The unit cell, the 200 plane and the 111 plane are outlined in black, green and red, respectively. **d,** Ca-mediated intermolecular

interactions between ferritin molecules in the lattice. Ca^{2+} ions (blue) are coordinated by two pairs of D84 and Q86 side-chains (magenta). **e,** Schematic representation of the formation, expansion and contraction of ferritin crystal-hydrogel hybrids.

8-hydroxypyrene-1,3,6-trisulfonic acid (pyranine) to the aforementioned co-monomer mixture. Pyranine has been reported to become covalently incorporated into the polymer backbone upon radical-mediated crosslinking and undergo a shift²⁸ in its emission maximum from 512 to 420 nm. Thus, the extent of in crystallo polymerization could be monitored through the decrease of green fluorescence intensity (emission wavelength $\lambda_{\text{emission}} = 500\text{--}550 \text{ nm}$, excitation wavelength $\lambda_{\text{excitation}} = 488 \text{ nm}$), indicating that hydrogel formation was complete in less than 2 min for a crystal with $l_{\text{edge}} = 70 \mu\text{m}$ (Extended Data Fig. 2b, c and Supplementary Video 1; see Extended Data Fig. 3 for polymer quantification via ^{19}F nuclear magnetic resonance, NMR). Polymerization was promptly followed by intrusion of the aqueous NaCl solution into the crystal-hydrogel matrix, which was clearly visualized owing to the difference between the refractive indices of the salt solution ($n_D = 1.3676$) and the matrix ($n_D \approx 1.34$) (Extended Data Fig. 2b and Supplementary Videos 1, 2). The solvent permeation process typically finished within 10 min and was accompanied by a small but noticeable enlargement of the crystals ($\leq 5\%$ increase in edge length) (Extended Data Fig. 2c).

Full expansion of hydrogel-infused ferritin crystals was initiated by placing them in deionized water. As observed using light microscopy, the expansion of the crystals was highly isotropic and their sharply faceted, polyhedral morphologies were preserved even after they grew to $\geq 210\%$ of their original dimensions (Fig. 2a and Supplementary Video 2; see Extended Data Fig. 4a and b for additional examples), often without the appearance of any defects. The expansion kinetics was biphasic, with time constants $\tau_{\text{fast}} < 100 \text{ s}$ and $\tau_{\text{slow}} \gg 300 \text{ s}$ (Fig. 2a). Isotropic growth continued indefinitely, until the edges of the materials were not discernible, but we typically stopped the process after $< 10 \text{ min}$, when considerable expansion had already occurred. No substantial release of ferritin molecules from the lattices was evident during the first 50 min of expansion (Extended Data Fig. 4c). Addition of a concentrated monovalent salt solution (NaCl or KCl) led to rapid dehydration and isotropic contraction of the expanded crystals to nearly their original size (Fig. 2a and Extended Data Fig. 5a). Recovery

of the original crystal dimensions could be achieved by further addition of CaCl_2 , owing to the ability of Ca^{2+} to both screen the negatively charged polymer backbone more effectively and to re-engage specific interactions between ferritin molecules. The same effect was observed with other divalent metal-ion salts (Extended Data Fig. 5b). The expansion-contraction cycle could be repeated at least eight times without apparent loss in amplitude and change in crystal morphology when a monovalent metal-salt solution was used to induce contraction (Extended Data Fig. 6). We observed that crystals contracted with CaCl_2 displayed considerably smaller expansion owing to the enhanced strength of the polymer network and protein-protein interactions. In control experiments, we examined other hydrogel formulations, including hydrogels that only contained polar but non-charged (pAm or poly-tris(hydroxymethyl)methyl(acrylamide)) or non-polar (poly-N-isopropylacrylamide) side-chains (Extended Data Fig. 7). All of these polymers led to either dissolution or disintegration of crystals after initiation of in crystallo polymerization, suggesting a lack of substantial interactions between the functional groups on these polymers and on the ferritin surface.

Interestingly, pAc hydrogels promoted isotropic expansion of the crystals in the absence of Am co-monomers (Extended Data Fig. 7), indicating that carboxylate side-chains are the primary mediators of interactions with ferritin molecules. By contrast, treatment of ferritin crystals with pre-formed pAc polymers, which cannot diffuse into the lattice, led to crystal dissolution upon transfer into water (Extended Data Fig. 7c). Together, these observations confirm that (i) there are extensive non-covalent interactions between ferritin molecules and the p(Ac-Am) hydrogel matrix that preserve the structural integrity of even highly expanded crystals, and (ii) the hydrogel matrix continuously and uniformly pervades the entire lattice, thus promoting cooperative transmission of any lattice deformations to enable isotropic expansion-contraction.

We investigated the expansion-related changes in the lattice arrangement of ferritin molecules using small-angle X-ray scattering (SAXS). Initial experiments entailed bulk measurements of a large number

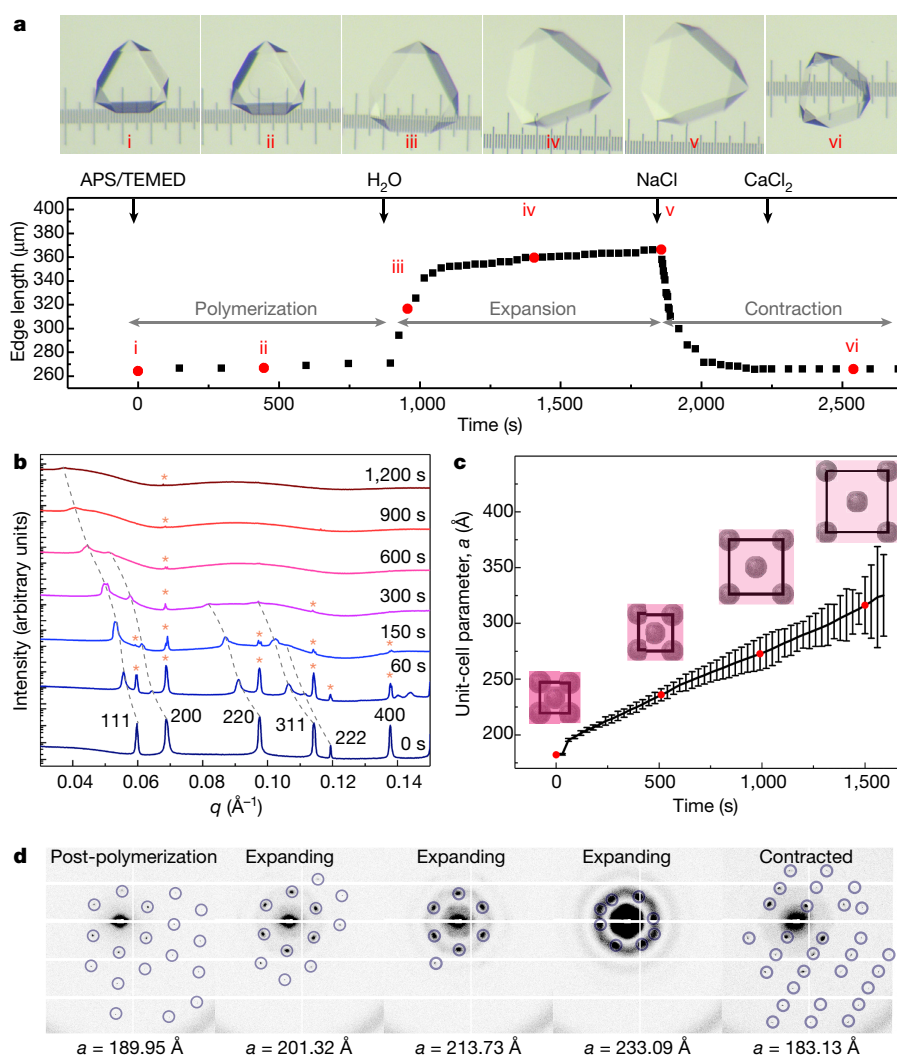


Fig. 2 | Characterization of the expansion and contraction behaviour of ferritin crystal-hydrogel hybrids. **a**, Structural evolution of a ferritin crystal-hydrogel hybrid during the polymerization-expansion-contraction process. Black arrows indicate the addition of different solutions or water to the crystal. The numbered images (i–vi) in the top panels correspond to the selected time points shown as red circles in the bottom panel. The separation between the major ticks of the ruler is 100 μm. **b**, SAXS profiles of hydrogel-infused ferritin crystals during lattice expansion, plotted against the scattering vector length q . The progression

of ferritin crystals suspended in a capillary tube. Figure 2b shows the evolution of the ‘powder’ SAXS pattern of more than 100 single p(Ac–Am)-infused crystals upon the initiation of polymerization though the addition of APS/TEMED in a solution that contains no salt; thus, polymerization is immediately followed by expansion (see Methods for experimental details). The spectrum of the unexpanded crystals is indicative of an fcc lattice with a unit-cell parameter of $a = 182.40$ Å. The isotropic growth of the unit cell is evident from the correlated shifts of the Bragg peaks to lower angles. The decay of the higher-angle peaks is considerably more rapid and is accompanied by the emergence of the ferritin form factor. This is consistent with the picture that as the crystal expands, the hydrogel matrix becomes less dense around the ferritin molecules, leading to their increased mobility. However, the (111) reflection is still evident after 20 min of expansion, which means that some long-range periodic order is still present when the unit cell has grown to $a = 325$ Å (Fig. 2c) and the volume of the material has increased to 570% of its original value.

To probe the reversibility of lattice expansion, we set up a microfluidic flow cell for single-crystal SAXS experiments (Extended Data Fig. 8), which circumvent the inherent issues associated with

of scattering peaks to lower angles is indicated with grey dashed lines. Peaks corresponding to the original lattice parameters (designated with red asterisks) are visible throughout the process. **c**, Changes in the unit-cell parameter a during lattice expansion, calculated from the SAXS profiles shown in **b**. The schematics correspond to the red circles and are drawn to scale. The error bars were determined from the full-widths at half-maximum of the scattering peaks. **d**, Expansion and contraction of a single crystal, monitored using SAXS.

bulk measurements in a small capillary tube (such as sample heterogeneity and inefficient solvent diffusion). The SAXS data in Fig. 2d indicate that a single-crystal lattice that has expanded by 27%—corresponding to a separation of 35 Å between neighbouring ferritin molecules—can return to its original dimensions upon NaCl/CaCl₂-induced contraction. To examine whether this recovery also occurs at the level of atomic periodicity, we conducted high-angle, single-crystal X-ray diffraction (XRD) experiments at room temperature (Fig. 3a–c). These experiments showed that crystals that expanded by up to 40% could fully regain their native diffraction pattern upon contraction with divalent metal-ion salts (Extended Data Fig. 5b). With such expanded and Ca-contracted crystals, we consistently obtained datasets with resolutions <1.15 Å at a synchrotron source at 100 K (Extended Data Table 1). Interestingly, the resulting crystal structures revealed two different conformational states of the Ca²⁺-bridged ferritin–ferritin interfaces (Fig. 3d): about 60% of these interfaces were found in the native configuration (as shown in Fig. 1d, but with a well resolved Ca-coordinated water molecule), whereas the remaining 40% presented an alternative coordination mode for Ca²⁺, probably stemming from lattice

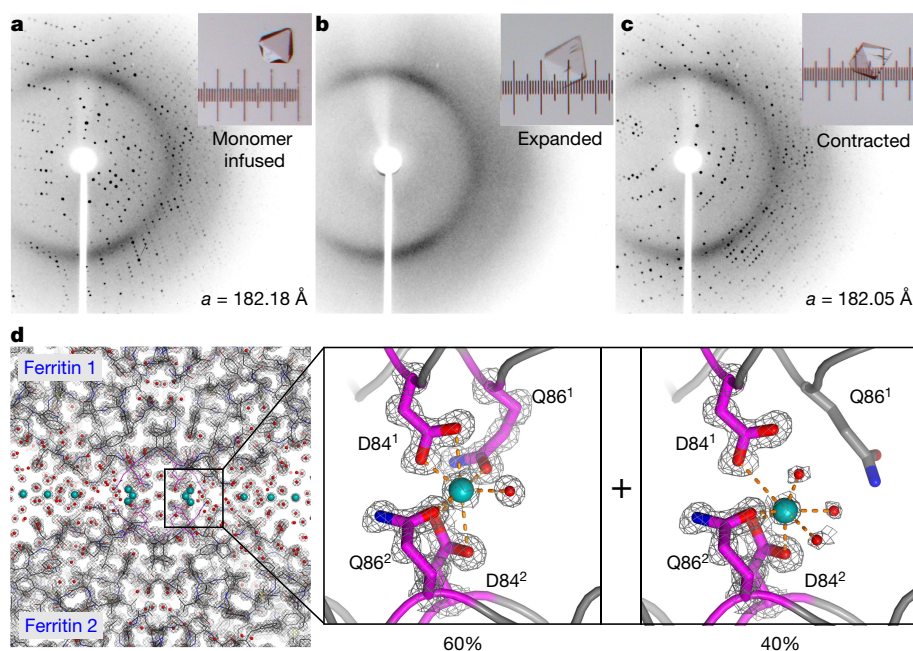


Fig. 3 | Atomic-level structural characterization of ferritin crystal-hydrogel hybrids by XRD. **a–c**, XRD patterns (at temperature $T = 293\text{ K}$) of a ferritin crystal infused with polymer precursors (**a**), after polymerization and expansion (**b**) and after contraction with CaCl_2 (**c**). Light micrographs of the crystal are shown in the insets; the separation between the major ticks of the ruler is $100\text{ }\mu\text{m}$. **d**, 1.06-Å-resolution

structure ($T = 100\text{ K}$; PDB ID, 6B8F) of the contracted ferritin crystal-hydrogel hybrid, showing the electron density surrounding the Ca-mediated ferritin-ferritin interfaces and highlighting the two observed Ca coordination conformations. The electron density ($2F_o - F_c$) map (grey) is contoured at 1.5σ . Water molecules and Ca ions are shown as red and blue spheres, respectively.

rearrangements during contraction. Notably, the 1.06-Å-resolution crystal structure (R -factors, $R_{\text{work}} = 9.10\%$; $R_{\text{free}} = 10.26\%$; estimated coordinate error (dispersion precision indicator), 0.011 Å) is the highest-resolution ferritin structure reported until now. Our findings suggest that hydrogel infusion and the expansion-contraction process do not diminish XRD data quality and may actually improve it.

Any local anisotropy developed during the expansion or contraction of the hydrogel matrix would be expected to cause dislocations in the embedded ferritin lattice. Indeed, exposure of hydrogel-infused crystals to rapid changes or temporary spatial gradients in NaCl or CaCl_2 concentrations frequently led to fracturing (see, for example, Fig. 3b inset). However, these materials showed a remarkable ability to self-heal, whereby the cracks were spontaneously and, in some cases, scarlessly sealed (Fig. 4a, b and Supplementary Video 3), owing to the reversible bonding interactions of the hydrogel network with the protein molecules (Fig. 4c). It is important to note that covalently crosslinked hydrogels like p(Ac-Am) do not typically self-heal unless they are modified with dynamic bonding functionalities^{22,29}. In the case of our materials, the role of such functional groups is fulfilled by the ferritin molecules, which act as interaction hubs for polymer chains. During expansion-contraction cycles, cracks tended to reoccur in the same loci in a given crystal (Supplementary Video 3). This observation suggests that the healed interfaces had not fully regained the original hydrogel crosslinking density of the bulk material, at least in the time scale (several minutes) of the experiments. Hydrogel integration substantially mitigated the brittleness of native ferritin crystals (Extended Data Fig. 9a). We observed no fragmentation, even in cases of substantial fracturing that propagated throughout the crystals, and fissures as wide as $20\text{ }\mu\text{m}$ could be closed to recover near-native crystal morphology (Fig. 4b). The ferritin crystal-hydrogel hybrids had a reduced modulus of about 1 GPa , which is similar to that of ferritin crystals (Extended Data Fig. 9b), but several orders of magnitude higher than those of hydrogels³⁰. The hybrids are also highly thermostable, maintaining their crystalline order at $\geq 80^\circ\text{C}$ (Extended Data Fig. 9c).

Owing to the inherent chemical tailorability of ferritin molecules, the crystal-hydrogel hybrids could be functionalized in different

ways. They could be constructed from ferritin molecules with mineralized ferrihydrite in their interior cavity (Extended Data Fig. 9d and Supplementary Video 4), thus exploiting ferritin's native function as a ferroxidase, or with fluorescent tags covalently attached to their exterior (Fig. 4d, e and Supplementary Video 4). Additionally, spatially differentiated, core-shell crystals were created using a layer-by-layer growth method (Fig. 4d, e). When infused with p(Ac-Am), such nanoparticle- or fluorophore-functionalized lattices displayed the same isotropic expansion-contraction behaviour as non-functionalized ones. The layer-by-layer growth process was further modified whereby the core lattice domain (labelled with rhodamine groups) was first covalently fixed through the chemical crosslinking of ferritin molecules with glutaraldehyde, followed by the growth of an uncrosslinked, unlabelled shell layer and the incorporation of the p(Ac-Am) polymer into the composite lattice. Hydration of such 'fixed core/expandable shell' crystals led to complete fragmentation of the shell layer due to the strain generated at the mechanically mismatched core-shell interface, exposing the morphologically unaltered core layer (Fig. 4e). These examples highlight the facility with which chemical and mechanical patterning are achieved in protein crystal-hydrogel hybrids.

We have reported here a new form of materials that integrate macromolecular protein crystals with synthetic polymer networks. These hybrids seamlessly combine the structural order and periodicity of crystals, the adaptiveness and tunable mechanical properties of polymeric networks and the chemical versatility of protein building blocks. Additionally, the ability to reversibly expand-contrast crystal lattices and mobilize their protein components may provide a new means to improve XRD quality and explore otherwise inaccessible protein structural states using three-dimensional protein crystallography. Protein crystals are often highly porous, sometimes containing up to 90% solvent, and are usually assembled through weak, non-covalent packing interactions; therefore, our approach should be applicable to other protein lattices. Their potential for generalizability, coupled with the chemical tailorability of synthetic polymers and the genetic mutability of proteins, should make protein crystal-hydrogel hybrids a rich medium for materials science.

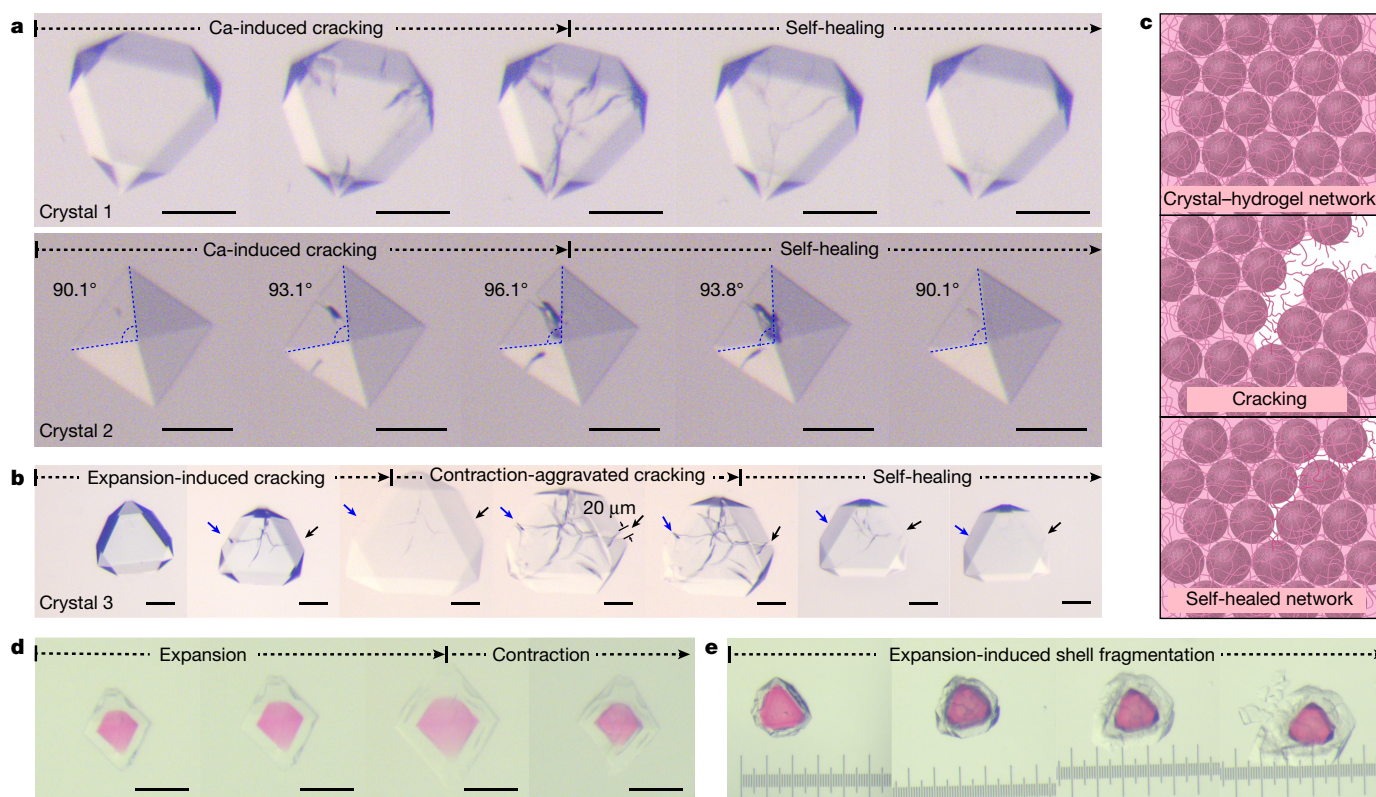


Fig. 4 | Self-healing behaviour and functionalization of ferritin crystal-hydrogel hybrids. **a**, Light microscopy images of crystal-hydrogel hybrids, showing the self-healing of cracks that appear during Ca-induced contraction. **b**, Extensive cracks can also appear during crystal expansion or during the initial stages of NaCl-induced contraction, but eventually self-heal. The arrows point to the termini of the major crack extending through the crystal. **c**, Schematic of crack formation and self-healing

through the interactions between polymer strands and ferritin molecules. **d**, Isotropic expansion and contraction of a crystal-hydrogel hybrid with an expandable core and an expandable shell. **e**, Swelling-induced fragmentation of the shell of a crystal-hydrogel hybrid with a fixed core and expandable shell. All scale bars, as well as the separation between the major ticks of the ruler in **e**, correspond to a length of 100 μm . See Supplementary Videos 3 and 4 that accompany this figure.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0057-7>.

Received: 9 October 2017; Accepted: 12 March 2018;

Published online 2 May 2018.

- Hawkins, T., Mirigian, M., Yasar, M. S. & Ross, J. L. Mechanics of microtubules. *J. Biomech.* **43**, 23–30 (2010).
- Fletcher, D. A. & Mullins, R. D. Cell mechanics and the cytoskeleton. *Nature* **463**, 485–592 (2010).
- Block, S. M., Blair, D. F. & Berg, H. C. Compliance of bacterial flagella measured with optical tweezers. *Nature* **338**, 514–518 (1989).
- Lewis, J. K., Bothner, B., Smith, T. J. & Siuzdak, G. Antiviral agent blocks breathing of the common cold virus. *Proc. Natl Acad. Sci. USA* **95**, 6774–6778 (1998).
- Lok, S.-M. et al. Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nat. Struct. Mol. Biol.* **15**, 312–317 (2008).
- Kobatake, S., Takami, S., Muto, H., Ishikawa, T. & Irie, M. Rapid and reversible shape changes of molecular crystals on photoirradiation. *Nature* **446**, 778–781 (2007).
- Kim, T., Al-Muhanna, M. K., Al-Suaidan, S. D., Al-Kaysi, R. O. & Bardeen, C. J. Photoinduced curling of organic molecular crystal nanowires. *Angew. Chem. Int. Edn* **52**, 6889–6893 (2013).
- Panda, M. K. et al. Spatially resolved analysis of short-range structure perturbations in a plastically bent molecular crystal. *Nat. Chem.* **7**, 65–72 (2015).
- Naumov, P., Chizhik, S., Panda, M. K., Nath, N. K. & Boldyreva, E. Mechanically responsive molecular crystals. *Chem. Rev.* **115**, 12440–12490 (2015).
- Barthelet, K., Marrot, J., Riou, D. & Ferey, G. A breathing hybrid organic-inorganic solid with very large pores and high magnetic characteristics. *Angew. Chem. Int. Edn Engl.* **41**, 281–284 (2002).
- Sakata, Y. et al. Shape-memory nanopores induced in coordination frameworks by crystal downsizing. *Science* **339**, 193–196 (2013).

- Rabone, J. et al. An adaptable peptide-based porous material. *Science* **329**, 1053–1057 (2010).
- Suzuki, Y. et al. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **533**, 369–373 (2016).
- Serre, C. et al. Role of solvent-host interactions that lead to very large swelling of hybrid frameworks. *Science* **315**, 1828–1831 (2007).
- Worthy, A. et al. Atomic resolution of structural changes in elastic crystals of copper(II) acetylacetonate. *Nat. Chem.* **10**, 65–69 (2017).
- Mason, J. A. et al. Methane storage in flexible metal-organic frameworks with intrinsic thermal management. *Nature* **527**, 357–361 (2015).
- Couck, S. et al. An amine-functionalized MIL-53 metal-organic framework with large separation power for CO_2 and CH_4 . *J. Am. Chem. Soc.* **131**, 6326–6327 (2009).
- Chen, Q. et al. A controllable gate effect in cobalt(II) organic frameworks by reversible structure transformations. *Angew. Chem. Int. Edn* **52**, 11550–11553 (2013).
- Ghosh, S. & Reddy, C. M. Elastic and bendable caffeine cocrystals: implications for the design of flexible organic materials. *Angew. Chem. Int. Edn* **51**, 10319–10323 (2012).
- Commins, P., Hara, H. & Naumov, P. Self-healing molecular crystals. *Angew. Chem. Int. Edn* **55**, 13028–13032 (2016).
- Tanaka, T. et al. Phase transitions in ionic gels. *Phys. Rev. Lett.* **45**, 1636–1639 (1980).
- Phadke, A. et al. Rapid self-healing hydrogels. *Proc. Natl Acad. Sci. USA* **109**, 4383–4388 (2012).
- Holtz, J. H. & Asher, S. A. Polymerized colloidal crystal hydrogel films as intelligent chemical sensing materials. *Nature* **389**, 829–832 (1997).
- Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
- Elliott, J. E., Macdonald, M., Nie, J. & Bowman, C. N. Structure and swelling of poly (acrylic acid) hydrogels: effect of pH, ionic strength, and dilution on the crosslinked polymer structure. *Polymer* **45**, 1503–1510 (2004).
- Theil, E. C. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu. Rev. Biochem.* **56**, 289–315 (1987).
- Lawson, D. M. et al. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–544 (1991).
- Kaya, D., Pekcan, Ö. & Yılmaz, Y. Direct test of the critical exponents at the sol-gel transition. *Phys. Rev. E* **69**, 016117 (2004).

29. Strandman, S. & Zhu, X. Self-healing supramolecular hydrogels based on reversible physical interactions. *Gels* **2**, 16 (2016).
30. Denisin, A. K. & Pruitt, B. L. Tuning the range of polyacrylamide gel stiffness for mechanobiology applications. *Appl. Mater. Interfaces* **8**, 21893–21902 (2016).

Acknowledgements We thank the following colleagues for assistance: R. Alberstein for data processing; A. Rheingold, C. Moore and M. Gembicky for XRD; S. Weigand, T. Weiss and I. Rajkovic for SAXS; W.-J. Rappel for confocal microscopy; A. Groisman for the design and assembly of microfluidic chips; Z. Hu for performing the nanoindentation experiments. This work was primarily funded by the US Department of Energy, DOE (BES, Division of Materials Sciences, Biomolecular Materials Program, DE-SC0003844 to F.A.T.). Additional funding was provided by NSF (DMR-1602537 to F.A.T. for SAXS studies). Crystallographic data were collected at Stanford Synchrotron Radiation Lightsource (SSRL) and the Crystallography Facility of the University of California, San Diego. SAXS data were collected at SSRL and the Advanced Photon Source. SSRL and the Advanced Photon Source are supported by the DOE Office of Science, Office of Basic Energy Sciences under contracts DE-AC02-76SF00515 and DE-AC02-06CH11357, respectively.

Author contributions L.Z., J.B.B. and F.A.T. conceived the work and designed the experiments. L.Z. conducted the majority of the experiments and performed light microscopy analysis. L.Z. and J.B.B. performed XRD and SAXS analyses. L.Z. and R.H.S. performed confocal microscopy analysis. F.A.T. wrote the manuscript with contributions from all authors.

Competing interests L.Z., J.B.B. and F.A.T. have submitted a patent application based on the work described here.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0057-7>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0057-7>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to F.A.T.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Protein expression, purification and characterization. The plasmid for the ΔC^* variant of human heavy-chain ferritin (HuHF), devoid of all native cysteine residues (C90E, C102A and C130A), was obtained via site-directed mutagenesis as previously described³¹. Expression and purification of ΔC^* was performed according to the previously published protocol³².

Determination of zeta potentials. Purified ferritin was concentrated to about 200 μ M and exchanged into a buffer solution containing 50 mM 2,2-Bis(hydroxymethyl)-2,2',2''-nitrilotriethanol (Bis-Tris) (pH 6.0), 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.0) or 50 mM HEPES (pH 7.5) using an Amicon Ultra centrifugal filter unit (10 kDa cutoff). The zeta potentials of ferritin in the three different buffers were determined using a Zetasizer Nano ZS90 (Malvern Instruments). Experimental runs were performed to collect 12 datasets with a He–Ne laser at 633 nm.

Formation of crystal–hydrogel hybrids. *Polymer precursor solution.* 25 mM (HEPES; pH 7.0), 30 mM CaCl_2 , 917 mM (8.625% w/v) sodium acrylate, 352 mM (2.5% w/v) acrylamide and 13 mM (0.2% w/v) N,N'-methylenebis(acrylamide).

Polymerization solution. 4 M NaCl, 1% (w/v) APS and 1% (v/v) TEMED. Octahedron-shaped ferritin crystals formed over 1–2 days in a buffered solution containing 25 mM HEPES (pH 7.0), 3–14.5 μ M protein (per 24meric ferritin cage) and 4.5–7.5 mM CaCl_2 . Once the ferritin crystals matured, the crystallization solution was replaced with the polymer precursor solution. Crystals were soaked for more than 10 h to ensure full infusion of the monomers into the ferritin crystals, and were then individually transferred with a mounted CryoLoop (Hampton) to the polymerization solution for 5 min, initiating in crystallo polymerization. Alternatively, the crystallization solution was replaced with the polymerization solution for the bulk polymerization of many crystals at once.

Measurement of the rate of diffusion into ferritin crystals. A large ferritin crystal was transferred with a mounted CryoLoop onto a glass slide, and 20 μ l of a solution containing 20 μ M (0.01 mg ml⁻¹) rhodamine B, 30 mM CaCl_2 and 25 mM HEPES (pH 7.0) was added to the crystal. The rhodamine diffusion process was monitored with a 10 \times air objective installed on a spinning-disk confocal Axio Observer inverted microscope (Zeiss) equipped with a pair of Quantum 5125C cameras (Roper), using a filter to collect light at 575–650 nm (red channel). Differential interference contrast (DIC) and fluorescence (564 nm excitation) images were captured at 1-s intervals with a 10-ms exposure. Images were collected in Slidebook 6 (Intelligent Imaging Innovations) and analysed using Fiji (<http://fiji.sc/Fiji>).

Determination of in crystallo polymerization dynamics. Ferritin crystals were incubated in a polymer precursor solution supplemented with 5.7 mM (0.3%) pyranine (Sigma–Aldrich). After 12 h, an individual crystal was transferred onto a glass slide and polymerization was initiated by adding 10 μ l of the polymerization solution. Hydrogel polymerization throughout the crystal and the corresponding decrease of pyranine fluorescence were monitored with a 20 \times air objective on the confocal microscope as described above, using a filter to collect light at 500–550 nm (green channel). DIC and fluorescence (488 nm excitation) images were captured at 1-s intervals with 100-ms (DIC) and 1-s (fluorescence) exposures.

Scanning electron microscopy of ferritin crystals. Native ferritin crystal and crystal–hydrogel hybrid samples were deposited onto glow-discharged, Formvar/carbon-coated Cu grids (Ted Pella Inc.). Each grid was blotted with filter paper to remove excess liquid. Grids were mounted onto a STEM 12x v2 sample holder and imaged using a Sigma 500 scanning electron microscope (Zeiss) at an accelerating voltage of 1 kV using a 30- μ m aperture.

Polymer quantification with ¹⁹F NMR. Large-scale crystallization of ferritin was carried out in a 24-well culture plate (Costar). 100 μ l of 25 μ M ferritin in 15 mM Tris (pH 7.4) and 150 mM NaCl was combined with 100 μ l of a buffered solution containing 50 mM HEPES (pH 7.0) and 12 mM CaCl_2 . Crystals formed overnight and matured over 72 h. The solution in each well was replaced with 100 μ l of a polymer precursor soaking solution containing: 25 mM HEPES (pH 7.0), 30 mM CaCl_2 , 179.9 mM 2-(trifluoromethyl)acrylic acid, 744.8 mM sodium acrylate, 350.7 mM acrylamide and 20.4 mM N,N'-methylenebis(acrylamide). After soaking overnight, this solution was removed, and the crystals were washed with a buffered solution (25 mM HEPES, pH 7.0; 30 mM CaCl_2) to remove unincorporated monomers. Polymerization was initiated by replacing the washing solution with 100 μ l of the polymerization solution. After 10 min, the crystals were transferred into an Eppendorf tube and centrifuged at 2,000g for 60 s. The supernatant was decanted, and the crystals were resuspended in 1 ml D₂O. Concentrated HCl was added until the pH of the solution was approximately 4.0 to facilitate crystal decomposition. 705 μ l of this solution was transferred into an NMR tube and supplemented with 4.6 mM trifluoroacetic acid. The ¹⁹F-NMR spectrum was collected using a 300M Bruker AVA spectrometer with a ¹⁹F probe (Extended Data Fig. 3). The peak at –64.94 p.p.m. corresponds to free 2-(trifluoromethyl)acrylic acid, the cluster of peaks near –67.07 p.p.m. to 2-(trifluoromethyl)acrylic acid that has been incorporated into the polymer, and the peak at –75.51 p.p.m. to the trifluoroacetic acid standard. From the integration of these peaks it was deduced

that (a) the total concentration of 2-(trifluoromethyl)acrylic acid in the solution was 8.2 mM and (b) 74.7% of the monomeric precursor was incorporated into the polymer matrix inside the crystals. The protein concentration was determined to be 60.0 μ M using the Bradford assay³³, and the molar ratio of 2-(trifluoromethyl)acrylic acid to ferritin was calculated as 137:1. Given this ratio and the fact that each unit cell of the ferritin crystals contains four ferritin cages and has a volume of about 5832 nm³ ($a = 18$ nm), the concentration of 2-(trifluoromethyl)acrylic acid in the crystal lattice was calculated as 155.6 mM, which is very similar to its concentration (179.9 mM) in the soaking solution.

Monitoring crystal expansion–contraction using light microscopy. Single crystals were transferred with a mounted CryoLoop onto a glass slide with a microscopic ruler (OMAX). All images and videos were obtained on an SZX7 (Olympus) microscope equipped with an Infinity 1 charge-coupled device (CCD; Lumenera). For crystals that had not been polymerized, 10 μ l of the polymerization solution was carefully added to minimize crystal movement. This solution was removed before water addition. For previously polymerized crystals, water (Milli-Q, 30 μ l) was added and crystal expansion was observed over 5–20 min. To initiate crystal contraction, water was replaced with a solution containing either 4 M NaCl or 1 M CaCl_2 . This expansion–contraction cycle could be repeated at least eight times for a crystal if NaCl was used to induce crystal contraction. Crystal size was determined by measuring the edge length of a facet relative to the microscopic ruler using the Fiji image processing package.

Monitoring crystal expansion using confocal microscopy. Crystals containing polymer precursors were prepared as described above. One of these crystals was transferred onto a glass slide and imaged on a confocal microscope. After capturing an initial image, the crystal was polymerized in 10 μ l of the polymerization solution, and its expansion in 30 μ l water was monitored. DIC images were captured at different time intervals with a 100-ms exposure until the crystal was no longer visible.

Quantification of protein release during expansion. Large-scale crystallization of ferritin was carried out as described above. Once crystals fully matured, the well solution was replaced with 100 μ l of the polymer precursor solution. After 12 h, the crystals were all combined into a single Eppendorf tube and 500 μ l of the polymerization solution was added. Crystals were expanded by replacing the polymerization solution with 1 ml water. During this experiment, aliquots (100 μ l) of the protein solution were removed and replaced with 100 μ l of water, and each aliquot was used to determine the protein concentration using the Bradford assay.

Multi-crystal expansion monitored using SAXS. Crystals for multi-crystal small-angle X-ray scattering were prepared as described above and transferred into the polymer precursor solution. A large number ($n > 100$) of crystals were transferred to an Eppendorf tube. After the crystals had settled at the bottom, they were transferred, along with 50 μ l of solution, into a 1.5-mm quartz capillary (Hampton). Crystals in capillaries were analysed at beamline 5-ID-D of the Advanced Photon Source (Argonne National Laboratory). Data were collected using collimated X-ray radiation (0.7293 Å, 17 keV) calibrated with both a glassy carbon standard and a silicon diffraction grating. After the sample was mounted on the instrument, a thin tube (with a diameter of 0.51 mm) was inserted into the capillary to facilitate the addition of 50 μ l of solution with a syringe injector during X-ray exposure. The injected solution contained a more concentrated polymerization solution without NaCl (2% APS and 2% TEMED) in water. After the first exposure, the solution was injected, and an image with a 1-s X-ray exposure was collected every 30 s. Peaks corresponding to the original lattice were visible throughout the process, indicating that some of the crystals in the bulk sample did not expand. This is probably due to limited solvent diffusion or incomplete polymerization within the capillary tubes used for the SAXS experiments. It is important to note that in this procedure, 'polymerized' crystals immediately began expanding upon the commencement of data collection. The reason for this experimental strategy (instead of polymerization in a high-ionic-strength solution, followed by the initiation of expansion through lowering the ionic strength) is that it was not possible to sufficiently dilute the high-ionic-strength polymerization solution in the thin capillary tubes used for SAXS (which cannot accommodate addition of large volumes of solution) to enable expansion.

Scattered radiation was detected using a CCD area detector and one-dimensional scattering data were obtained through the azimuthal averaging of the two-dimensional data to produce plots of the scattering intensity as a function of the scattering vector length, $q = 4\pi \sin(\theta/\lambda)$, where θ is one-half of the scattering angle and λ is the wavelength of the X-rays used. Analysis of the one-dimensional data was performed using the powder diffraction processing software JADE (MDI) or Origin (OriginLab).

Multi-crystal SAXS at elevated temperatures. Large-scale crystallization of ferritin was performed as described above. The crystallization solution was removed, and ferritin crystals were resuspended in either the polymer precursor solution or a buffered solution containing 25 mM HEPES (pH 7.0) and 30 mM CaCl_2 . After 72 h, the polymer precursor soaking solution was replaced with the polymerization solution. After 10 min, this was also replaced with a buffered solution containing

25 mM HEPES (pH 7.0), 4 M NaCl and 30 mM CaCl_2 . Both samples, containing either native ferritin crystals or the crystal–hydrogel hybrids, were transferred into 1.5-mm quartz capillaries (Hampton). Data were collected at beamline 4-2 of SSRL using collimated X-ray radiation (1.1271 Å, 11 keV) calibrated with a silver behenate standard. The samples were heated using a custom-built thermal stage operating at 1°C min^{-1} , and images with a 1-s X-ray exposure were collected every minute. Scattered radiation was detected using a Pilatus3 X 1M detector (Detectors) and processed as described above.

Single-crystal SAXS. Crystals containing the polymer precursors were prepared and polymerized using the polymerization solution as described above, and were analysed at SSRL (beamline 4-2). Single crystals were harvested with a mounted CryoLoop and transferred into a 2 M NaCl solution in the 400- μm -diameter central well of a custom-made microfluidic chip (Extended Data Fig. 8a, b). The microfluidic chip was sealed with a coverslip, attached to a syringe injector and mounted on beamline 4-2 at SSRL for data collection. Data were collected using collimated X-ray radiation (1.127 Å, 11 keV) calibrated with a silver behenate standard. Water was injected into the microfluidic chip at $1\ \mu\text{l s}^{-1}$ to initiate expansion, and 0.5-s-exposure images were taken every about 2.5 s for 4 min. After the data acquisition for crystal expansion was complete, the process was repeated—in the order 4 M NaCl, water, 1 M CaCl_2 , water—to monitor repeated contraction and expansion processes. Data were collected using a Pilatus3 X 1M detector (Detectors). The unit-cell parameters were determined by calculating the radial distance of individual reflections, after fitting the spot intensity to a two-dimensional Gaussian surface.

Single-crystal XRD at room temperature. Crystals containing polymer precursors were prepared and imaged using light microscopy as described above. A single crystal was transferred onto a MicroMount precision tool (MiTeGen) with a 100- μm aperture and sealed with a MicroRT capillary (MiTeGen). Data were acquired on an APEX II CCD diffractometer (Bruker) using $\text{Cu K}\alpha$ radiation (1.5418 Å) at 295 K. Three images (60-s exposure) were collected at rotation angles $\varphi = 0^\circ, 60^\circ$ and 120° . The crystal was removed from the instrument and soaked in $10\ \mu\text{l}$ of the polymerization solution for 2 min. The crystal was transferred onto a microscopic ruler and $30\ \mu\text{l}$ of water was added. Crystal expansion was measured over 3 min. This crystal was returned to the MicroMount with the MicroRT capillary and an identical three-image dataset was collected. This process was repeated using $30\ \mu\text{l}$ of a 1 M CaCl_2 solution. After the crystal had contracted (1 min), another three-image dataset was collected. Images were analysed with the Apex III software (Bruker).

Single-crystal XRD at 100 K. Crystal–hydrogel hybrids were prepared and imaged using light microscopy as described above. Two crystals were harvested, $30\ \mu\text{l}$ of water was added, and crystal expansion was monitored over 5 min for both crystals. After 5 min, the water was removed and $30\ \mu\text{l}$ of either a solution containing 1 M CaCl_2 (crystal A) or 4 M NaCl (crystal B) was added. Crystal B was re-expanded in $30\ \mu\text{l}$ water. After 5 min, the water was replaced with $30\ \mu\text{l}$ of a 1 M CaCl_2 solution to contract crystal B. After contraction, both crystals were cryoprotected in perfluoropolyether (Hampton) and frozen in liquid N_2 . Single-crystal XRD data for the contracted ferritin crystals were collected at 100 K at beamline 9-2 of SSRL using 0.98-Å radiation. The data were integrated using iMosflm³⁴ and scaled with Aimless³⁵ (Extended Data Table 1). The structures for crystal A and crystal B were determined at resolutions of 1.06 Å and 1.13 Å, respectively. Molecular replacement was performed with Phaser³⁶ using the HuHF structure (PDB ID, 5CMQ) as a search model. Rigid-body, positional, anisotropic thermal and atom-occupancy refinements were carried out using Phenix³⁷. Coot³⁸ was used for iterative manual model building. The interstitial solvent content was calculated by subtracting the solvent volume of each crystal from the volume of the inner cavity of ferritin (calculated using VOIDOO)³⁹. All figures were produced with Pymol⁴⁰.

Nanoindentation measurements of crystals. The mechanical properties of the native ferritin crystals and the crystal–hydrogel hybrids were determined using a Hysitron TI 950 Triboindenter test instrument (Bruker). All crystals were dried before the indentation experiments. A Berkovich probe (TI-0039, 142.3° , 100 nm tip radius) was used to determine the hardness and reduced modulus of the native crystals and crystal–hydrogel hybrids. Experiments were conducted in displacement control mode using a displacement of 1,000 nm.

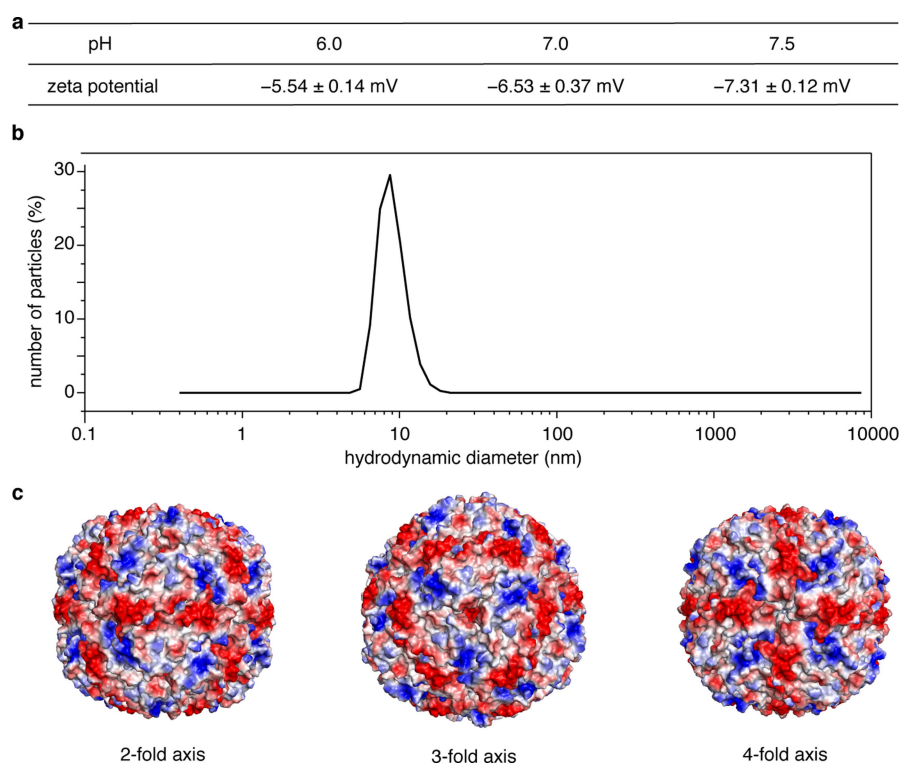
Preparation of iron-loaded ferritin. Iron-loaded ferritin was prepared by adding 10.8 ml of 10 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ over 2 h to 144.8 ml of a vigorously stirring solution containing 1 μM ferritin, 15 mM Tris (pH 7.4) and 150 mM NaCl. Subsequently, the solution was stirred for an additional hour before being concentrated to about 3 ml using a 10-kDa Amicon membrane. A 10DG column (Bio-Rad) was used to remove any unbound iron. The iron content was assessed using a 2,2'-bipyridine-based colorimetric assay⁴¹ and the protein concentration was determined using the Bradford assay. Each ferritin cage contained about 800 Fe atoms.

Formation of core-shell ferritin crystals. *Expandable core/expandable shell crystals.* Mature ferritin crystals were transferred to a buffered solution containing 25 mM HEPES (pH 7.0), 30 mM CaCl_2 and 1.9 mM $(1\ \text{mg ml}^{-1})$ 5-(and 6)-carboxytetramethyl rhodamine succinimidyl ester (NHS-rhodamine; Thermo Fisher Scientific). After soaking for 12 h, an individual crystal was removed and washed three times in a buffered solution containing 25 mM HEPES (pH 7.0) and 30 mM CaCl_2 to remove unbound NHS-rhodamine. The crystal was transferred to a well containing $10\ \mu\text{l}$ of 12.5 μM ferritin, 25 mM HEPES (pH 7.0) and 6 mM CaCl_2 . A transparent layer of ferritin formed around the rhodamine-labelled ferritin crystal over 12 h (creating a red core and a transparent shell). This crystal was soaked in a polymer precursor solution and polymerized as described above to yield an expandable core/expandable shell crystal.

Fixed core/expandable shell crystals. Fixed core/expandable shell ferritin crystals were prepared similarly to the expandable core/expandable shell crystals described above. The only difference was that after the rhodamine labelling step, the crystal was transferred into a solution containing 2.5% (v/v) glutaraldehyde, 25 mM HEPES (pH 7.0) and 30 mM CaCl_2 . After 12 h, the crystal was washed five times with water to remove unbound glutaraldehyde, followed by the epitaxial growth of the transparent layer of ferritin crystals on top of the core layer in a fresh crystallization solution containing 12.5 μM ferritin. This crystal was then soaked in a polymer precursor solution and polymerized as described above to yield a fixed core/expandable shell crystal.

Data availability. Crystal structures have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank under accession codes 6B8F (ferritin–polymer hybrid crystal 1; <https://www.rcsb.org/structure/6b8f>) and 6B8G (ferritin–polymer hybrid crystal 2; <https://www.rcsb.org/structure/6b8g>). Raw data are available from F.A.T.

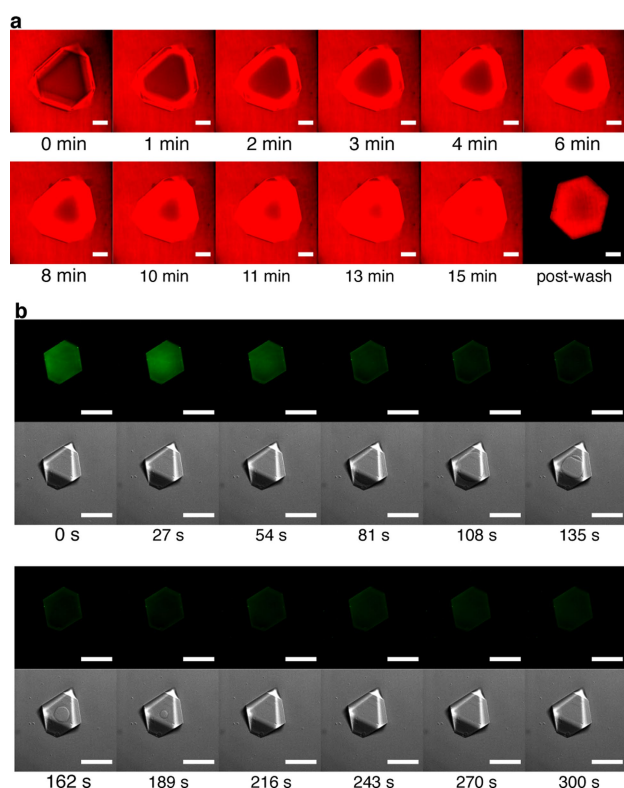
- Huard, D. J., Kane, K. M. & Tezcan, F. A. Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nat. Chem. Biol.* **9**, 169–176 (2013).
- Sontz, P. A., Bailey, J. B., Ahn, S. & Tezcan, F. A. A metal organic framework with spherical protein nodes: rational chemical design of 3D protein crystals. *J. Am. Chem. Soc.* **137**, 11598–11601 (2015).
- Bradford, M. M. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281 (2011).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).
- The PyMOL Molecular Graphics System Version 1.3, <https://pymol.org/2/support.html> (Schrödinger LLC).
- Levi, S. et al. Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. *J. Biol. Chem.* **263**, 18086–18092 (1988).



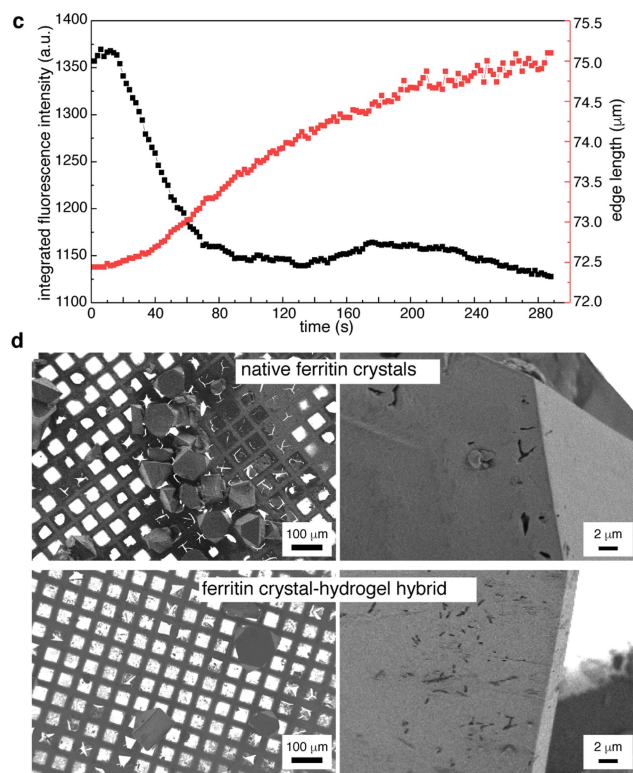
Extended Data Fig. 1 | Distribution of electrostatic charge on the surface of ferritin and size distribution of ferritin in solution.

a, pH-dependent zeta potentials of ferritin, determined by dynamic-light-scattering measurements. **b**, Dynamic-light-scattering profile of ferritin (200 μ M) in a solution of 50 mM HEPES (pH 7.0). **c**, Representation of the

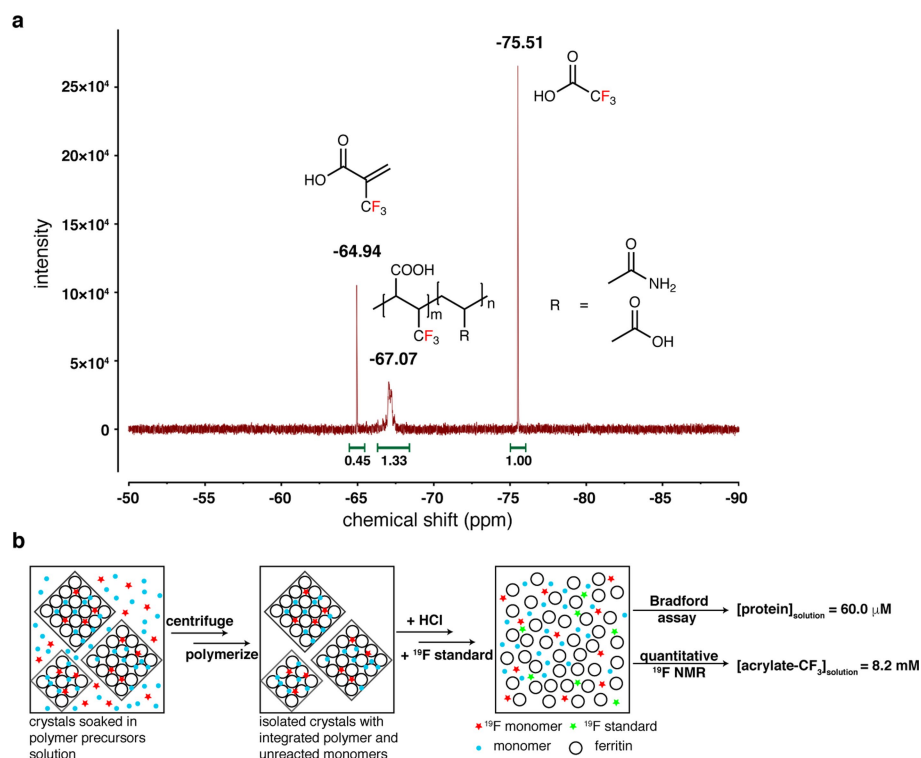
electrostatic charge distribution on the ferritin surface, as viewed along the two-, three- and four-fold symmetry axes. Positive ($+5k_B T/e$) and negative ($-5k_B T/e$) charges are shown in blue and red, respectively. k_B , Boltzmann constant; e , electron charge.



Extended Data Fig. 2 | Molecular diffusion and polymerization in ferritin crystals, monitored using confocal microscopy. **a**, Diffusion of rhodamine B into a ferritin crystal over 15 min. **b**, **c**, In crystallo polymerization of the hydrogel network, monitored through the decrease of integrated pyranine fluorescence (green fluorescence channel). The corresponding bright-field (DIC) images show the diffusion of the



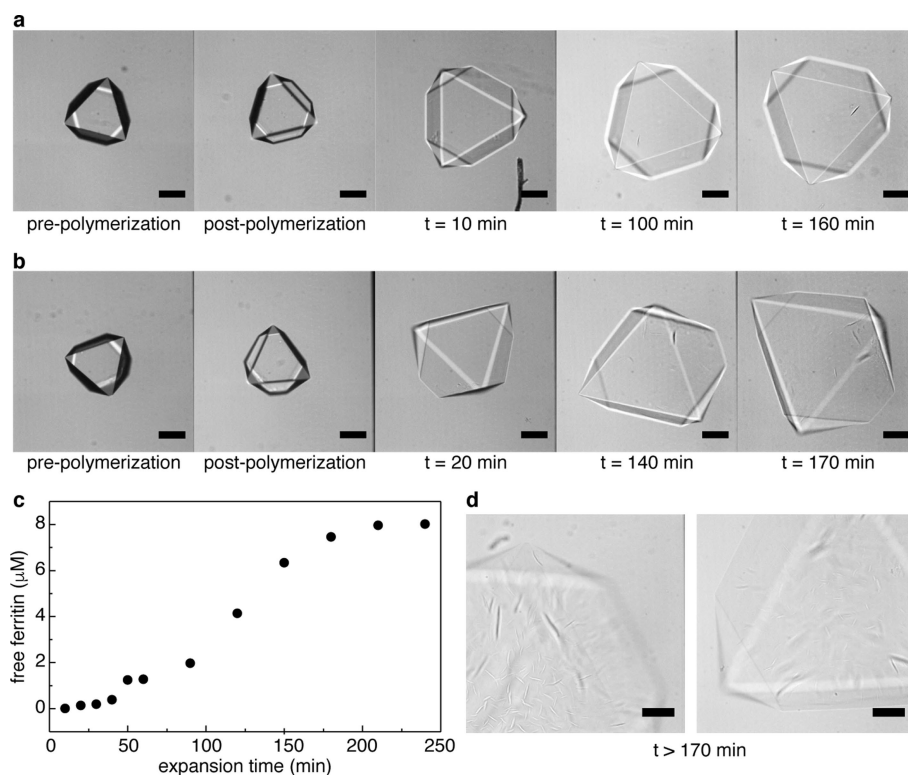
aqueous NaCl solution into the crystal. The ring-shaped diffusion front becomes evident at time $t = 108$ s and disappears by $t = 216$ s. The crystal expands by approximately 5% (edge length) during polymerization. Scale bars in **a** and **b** correspond to 100 μm . **d**, Scanning electron microscopy images of native ferritin crystals (top) and crystal-hydrogel hybrids (bottom).



Extended Data Fig. 3 | Quantification of an acrylic acid analogue using ^{19}F NMR. **a**, ^{19}F -NMR spectrum, showing peak assignments for the trifluoroacetic acid standard, free 2-(trifluoromethyl)acrylic acid, and 2-(trifluoromethyl)acrylic acid incorporated into the polymer.

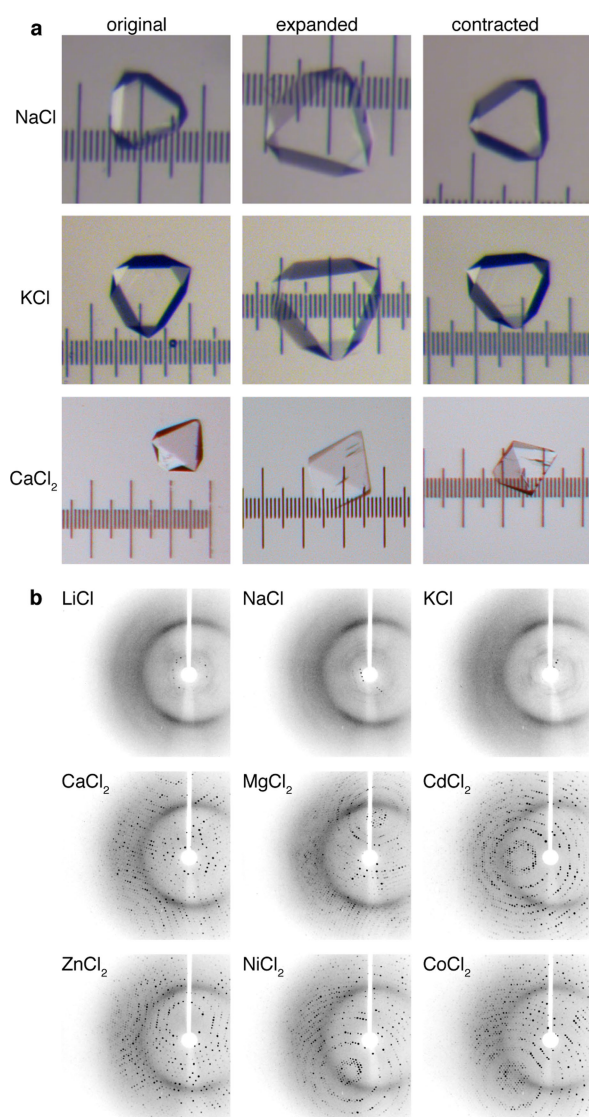
b, Diagram illustrating the experimental protocol for the quantification

of 2-(trifluoromethyl)acrylic acid uptake into the crystal lattice. The concentration of the 2-(trifluoromethyl)acrylic acid in the crystal lattice (155.6 mM) is approximately the same as its concentration in the soaking solution (see Methods for details).

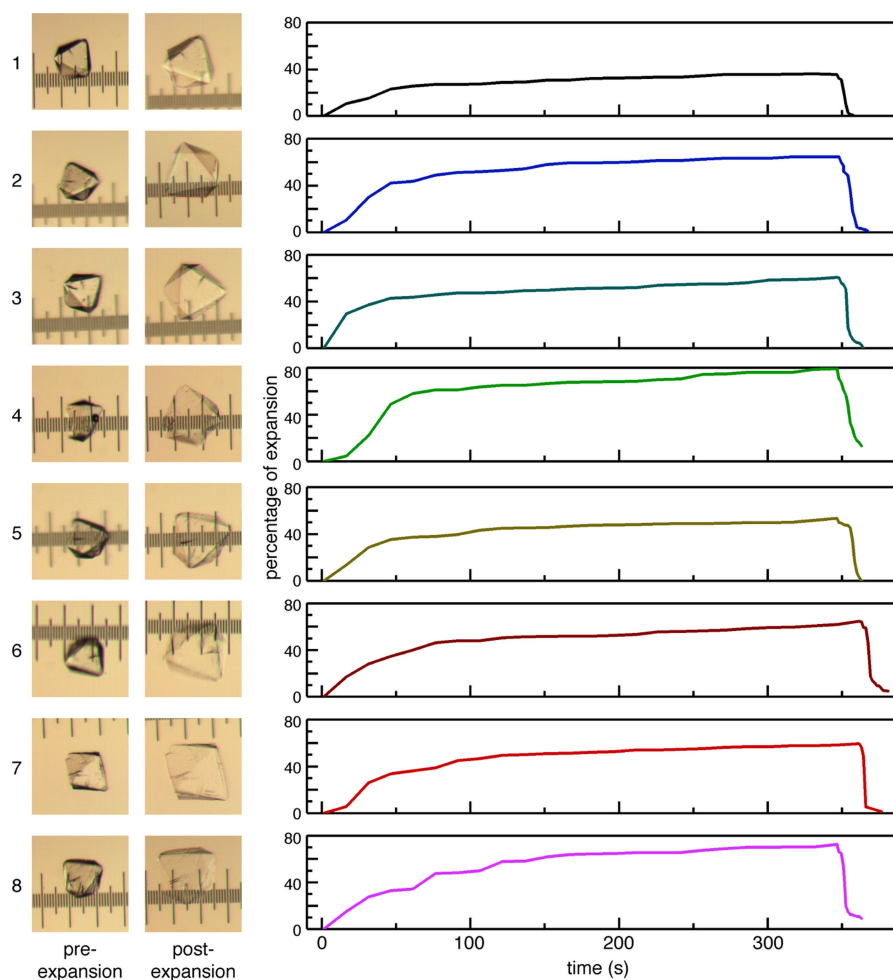


Extended Data Fig. 4 | Isotropic hyperexpansion of ferritin crystal-hydrogel hybrids. **a, b**, Continuous expansion of two different crystal-hydrogel hybrids in deionized water, monitored using confocal microscopy. Crystal facets are still discernible after expansion for more than 2 h. Scale bars correspond to $100\mu\text{m}$. **c**, Ferritin release into the

solution from expanding crystal-hydrogel hybrids ($n > 10,000$) over about 4 h. Negligible ferritin release is observed until about 1 h. Protein concentrations were determined using the Bradford assay. **d**, Confocal microscopy images of highly expanded crystal-hydrogel hybrids, showing the structural deterioration of the facets and the edges.

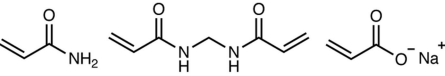
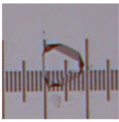

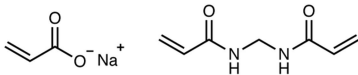
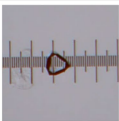
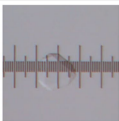
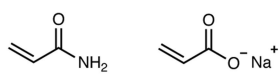


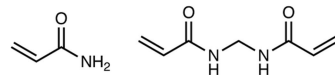
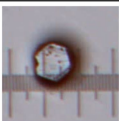
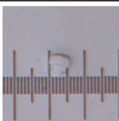
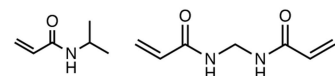
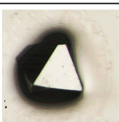
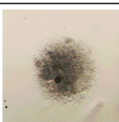
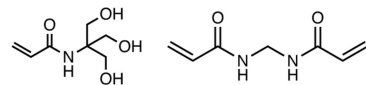
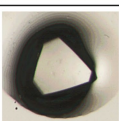

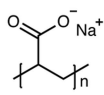
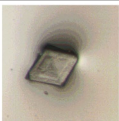
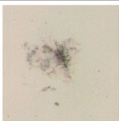


Extended Data Fig. 5 | Expansion and contraction behaviour of crystal-hydrogel hybrids in the presence of different metal ions. a, Light micrographs of the crystal-hydrogel hybrids at different stages of expansion and contraction in response to different metal ions. **b**, XRD patterns ($T = 273$ K) of expanded crystal-hydrogel hybrids, acquired upon contraction with different metal ions. Contraction with divalent cations (Ca, Mg, Cd, Zn, Ni and Co) reproducibly leads to the recovery of the full atomic-level order, whereas contraction with monovalent cations (Li, Na and K) only reinstates low-order diffraction peaks.



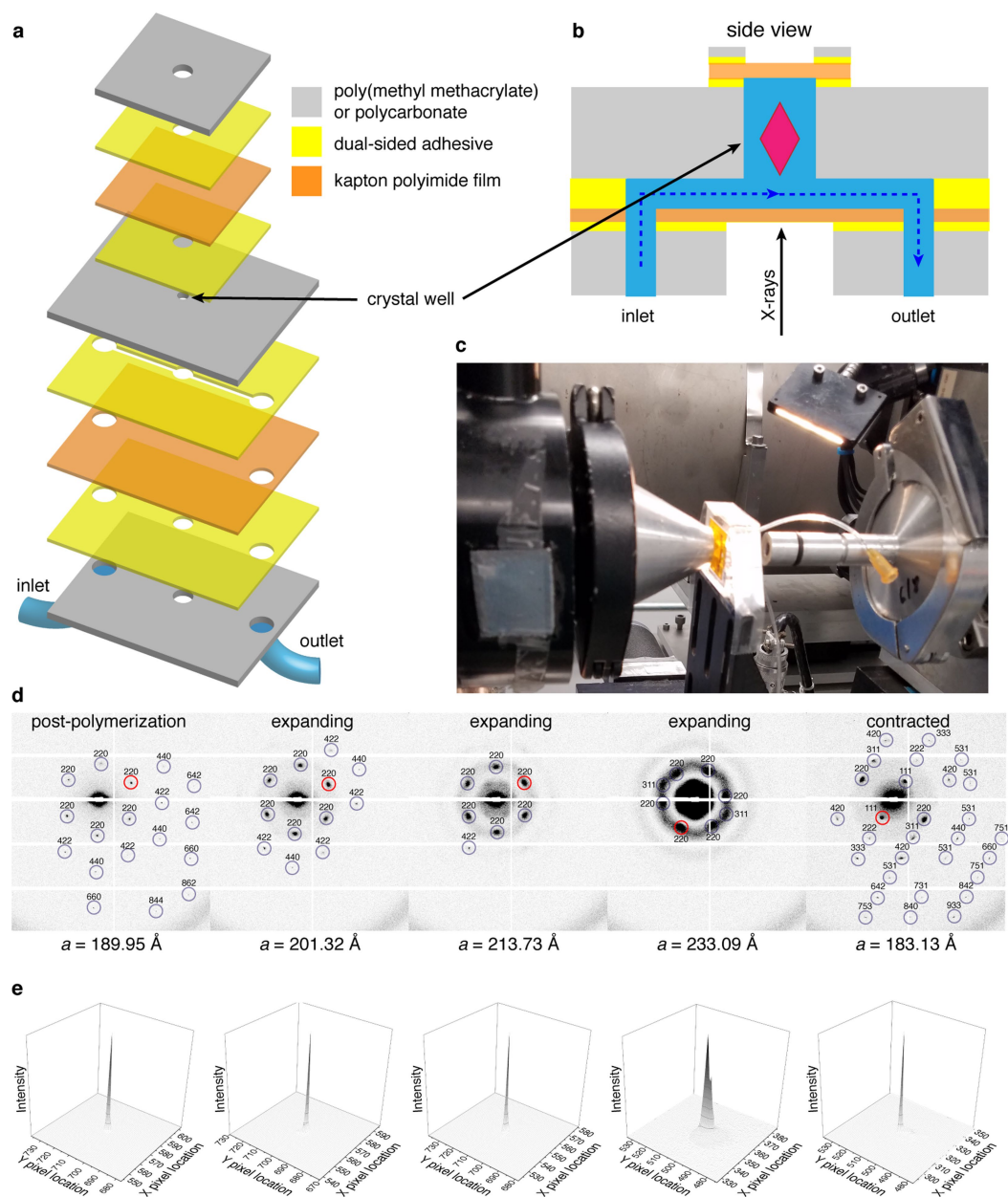
Extended Data Fig. 6 | Successive expansion-contraction cycles for a single ferritin crystal-hydrogel hybrid. Light micrographs of a hybrid crystal at pre- and post-expansion stages in each cycle are shown on the left, and the corresponding changes in edge length upon expansion-contraction are shown on the right. The separation between the major ticks of the ruler is $100\ \mu\text{m}$. The crystal expands to a lesser extent during

the first expansion cycle, which we ascribe to residual CaCl_2 (which forms strong polymer-polymer and protein-protein interactions) remaining in the solution that is transferred on the loop along with the crystal. The subsequent variability in the rate and extent of expansion is attributed to the different amounts of residual NaCl transferred in each cycle.

a					
polymer precursors	chemical structure		post-polymerization	expanded	
-sodium acrylate -MBAm -acrylamide					
-sodium acrylate -MBAm					
-sodium acrylate -acrylamide					
b					
polymer precursors	chemical structure		monomer infused	post-polymerization	
-acrylamide -MBAm					
-N-isopropylacrylamide -MBAm					
-TRIS-acrylamide -MBAm					
c					
polymer	chemical structure		polymer soaked	hydration	
-sodium polyacrylate $M_w = \text{ca. } 2100 \text{ Da}$					

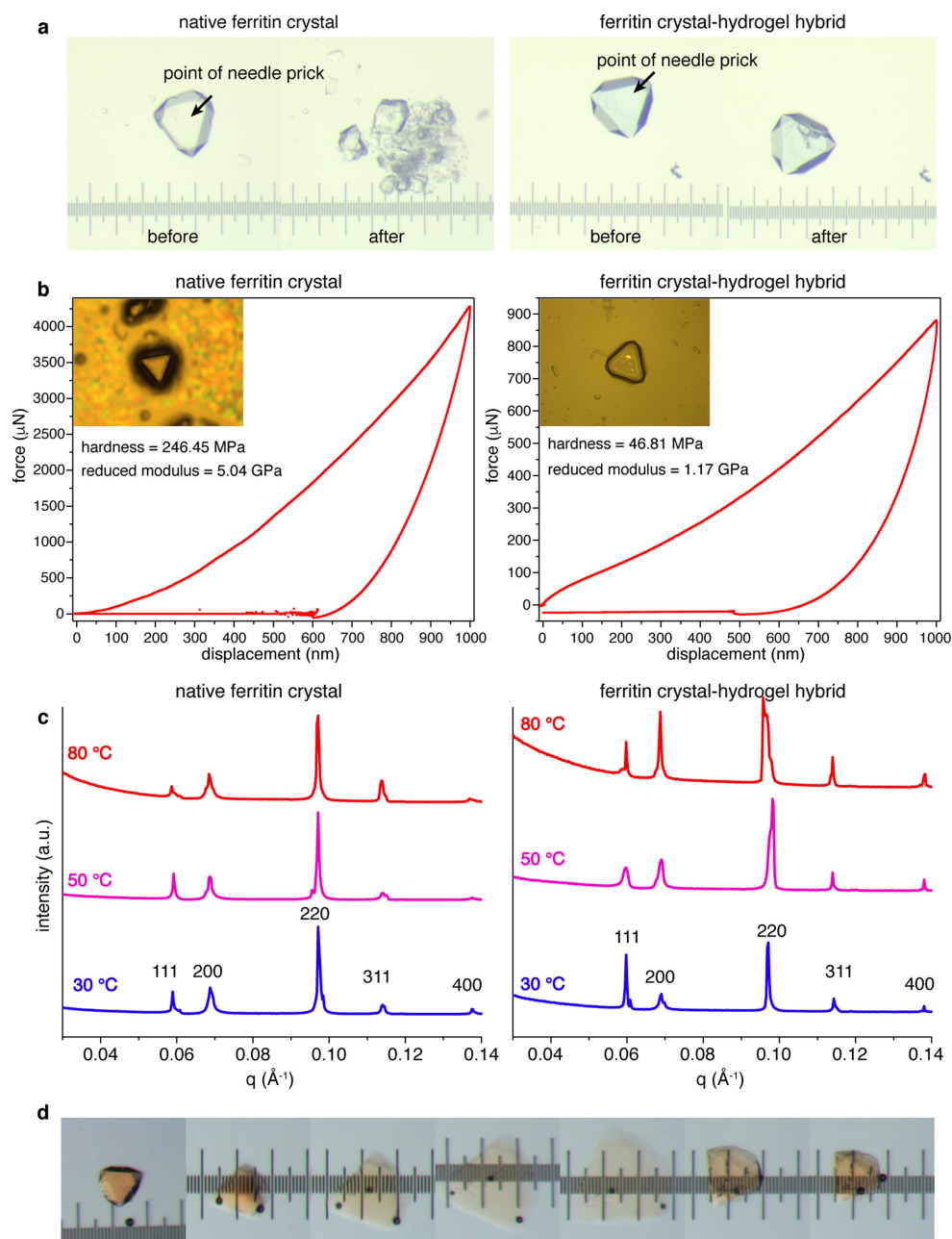
Extended Data Fig. 7 | Alternative hydrogel formulations. **a**, Alternative monomer combinations that yield successful in crystallo polymerization and crystal expansion. **b**, Monomer combinations that lead to crystal dissolution during polymerization. **c**, A crystal soaked in a solution

containing polyacrylate (molecular weight, $M_w = 2,100 \text{ Da}$) dissolves upon being transferred into water. The separation between the major ticks of the ruler is $100 \mu\text{m}$. MBAm, N,N'-methylenebis(acrylamide).



Extended Data Fig. 8 | SAXS imaging of a single crystal-hydrogel hybrid in a microfluidic chip. a, Schematic diagram of the microfluidic chip. **b,** Side-view representations of the microfluidic chip. **c,** Photograph of the microfluidic chip, mounted on beamline 4-2 at SSRL. **d,** Single-

crystal SAXS diffraction patterns observed at different stages of crystal expansion and contraction. The Miller indices of each visible spot are indicated. Reflections with the highest signal-to-noise ratio ($I/\sigma I$) are circled in red. **e,** Spot profiles of the highest- $I/\sigma I$ reflections indicated in **d**.



Extended Data Fig. 9 | Mechanical and thermal properties of ferritin crystal-hydrogel hybrids. **a**, Light-microscopy images showing the fragmentation of a native ferritin crystal and of a crystal-hydrogel hybrid upon application of external force with a needle at the location indicated with the arrow. The separation between the major ticks of the ruler is 100 μm . **b**, Temperature dependence of the SAXS profiles of native ferritin crystals and crystal-hydrogel hybrids. The small-angle reflections (that

is, periodic order) in both samples are maintained at 80 °C (the maximal temperature experimentally attainable). **c**, Determination of the hardness and reduced modulus of native ferritin crystals and crystal-hydrogel hybrids using atomic force microscopy nanoindentation measurements. **d**, Light-microscopy images showing the expansion and contraction of a crystal-hydrogel hybrid containing Fe-loaded ferritin molecules. The separation between the major ticks of the ruler is 100 μm .

Extended Data Table 1 | X-ray data collection and refinement statistics

	Crystal A (6B8F)	Crystal B (6B8G)
Data collection		
Space group	<i>F</i> 432	<i>F</i> 432
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	180.40	179.95
α , β , γ (°)	90	90
Resolution (Å)	63.65–1.06	63.62–1.13
Number of unique reflections	111189	92912
Multiplicity	14.0 (2.4)	32.2 (13.8)
<i>CC</i> 1/2	0.999 (0.897)	0.999 (0.628)
<i>R</i> _{merge}	0.060 (0.222)	0.131 (1.131)
$\langle I / \sigma \rangle$	24.8 (3.2)	19.1 (2.3)
Completeness (%)	99.0 (86.4)	100 (100)
Average mosaicity (°)	0.17	0.31
Total solvent content (%)	57.42	57.36
Interstitial solvent content (%)	39.72	39.62
Refinement		
<i>R</i> _{work} / <i>R</i> _{free}	0.0910/0.1026	0.1029/0.1213
Number of atoms		
Protein	1687	1699
Ligand/ion	13	13
Water	340	372
<i>B</i> -factors (Å ²)		
Protein	8.51	9.81
Ligand/ion	10.19	11.96
Water	21.83	23.75
Root-mean-square deviations		
Bond lengths (Å)	0.013	0.011
Bond angles (°)	1.317	1.241
MolProbity ⁴⁴ score	1.20	1.38
Clashscore	4.12	6.99
Ramachandran plot (%)		
Favoured	98.82	98.82
Outliers	0.00	0.00
Rotamers (%)		
Favoured	97.35	96.88
Poor	0.53	0.52
DPI (Å) ⁴⁵	0.011	0.015

Numbers in parentheses correspond to the highest resolution shell. *R*_{merge} and *CC* 1/2 are measurements used to determine an appropriate high-resolution limit for XRD data.

Synthesis, structure and reaction chemistry of a nucleophilic aluminyl anion

Jamie Hicks¹, Petra Vasko^{1,2}, Jose M. Goicoechea^{1*} & Simon Aldridge^{1*}

The reactivity of aluminium compounds is dominated by their electron deficiency and consequent electrophilicity; these compounds are archetypal Lewis acids (electron-pair acceptors). The main industrial roles of aluminium, and classical methods of synthesizing aluminium–element bonds (for example, hydroalumination and metathesis), draw on the electron deficiency of species of the type AlR_3 and AlCl_3 ^{1,2}. Whereas aluminates, $[\text{AlR}_4]^-$, are well known, the idea of reversing polarity and using an aluminium reagent as the nucleophilic partner in bond-forming substitution reactions is unprecedented, owing to the fact that low-valent aluminium anions analogous to nitrogen-, carbon- and boron-centred reagents of the types $[\text{NX}_2]^-$, $[\text{CX}_3]^-$ and $[\text{BX}_2]^-$ are unknown^{3–5}. Aluminium compounds in the +1 oxidation state are known, but are thermodynamically unstable with respect to disproportionation. Compounds of this type are typically oligomeric^{6–8}, although monomeric systems that possess a metal-centred lone pair, such as $\text{Al}(\text{Nacnac})^{\text{Dipp}}$ (where $(\text{Nacnac})^{\text{Dipp}} = (\text{NDippCR})_2\text{CH}$ and $\text{R} = \text{tBu, Me}$; $\text{Dipp} = 2,6\text{-iPr}_2\text{C}_6\text{H}_3$), have also been reported^{9,10}. Coordination of these species, and also of $(\eta^5\text{-C}_5\text{Me}_5)\text{Al}$, to a range of Lewis acids has been observed^{11–13}, but their primary mode of reactivity involves facile oxidative addition to generate $\text{Al}(\text{III})$ species^{6–8,14–16}. Here we report the synthesis, structure and reaction chemistry of an anionic aluminium(I) nucleophile, the dimethylxanthene-stabilized potassium aluminyl $[\text{K}\{\text{Al}(\text{NON})\}]_2$ ($\text{NON} = 4,5\text{-bis}(2,6\text{-diisopropylanilido})\text{-}2,7\text{-di-tert-butyl-}9,9\text{-dimethylxanthene}$). This species displays unprecedented reactivity in the formation of aluminium–element covalent bonds and in the C–H oxidative addition of benzene, suggesting that it could find further use in both metal–carbon and metal–metal bond-forming reactions.

Computational studies suggest that, relative to the corresponding boryl anion, heterocyclic (diamido)aluminyl systems of the type $[\text{Al}(\text{NRCH})_2]^-$ should have a higher energetic separation between the singlet ground electronic state and the triplet excited state (ΔE_{st}), and a lower energy associated with the group-13-centred lone pair^{9,17,18}. Despite this, no such systems have been experimentally realized. By making use of the steric demands and σ -electron-withdrawing properties of bulky arylamido substituents, together with a flexible chelating dimethylxanthene backbone, we have observed that aluminyl compounds can be synthesized that are stable up to 300 K, amenable to structural characterization in the solid state, and function as a nucleophilic source of aluminium in substitution chemistry.

Drawing inspiration from the two major classes of anionic carbon-centred nucleophiles ubiquitous in organic synthesis (namely, group 1 metal alkyl/aryl compounds and Grignard reagents)^{4,19}, we set out to synthesize potassium and magnesium aluminyl compounds of the types $[\text{K}\{\text{Al}(\text{NR}_2)_2\}]_n$ and $[\text{RMg}\{\text{Al}(\text{NR}_2)_2\}]_n$. Deprotonation of the bifunctional dimethylxanthene-derived secondary aniline $(\text{NON})\text{H}_2$ with $\text{K}[\text{N}(\text{SiMe}_3)_2]$, followed by reaction with AlI_3 , generates the $\text{Al}(\text{III})$ iodide $(\text{NON})\text{AlI}$ (Fig. 1, Extended Data Fig. 1 and Extended Data Table 1). $(\text{NON})\text{AlI}$ is a versatile substrate for reduction chemistry:

reaction with potassium graphite in toluene or benzene solution provides access to $\text{Al}(\text{I})$ or $\text{Al}(\text{II})$ products depending on reaction stoichiometry. The use of excess KC_8 forms the dimeric potassium aluminyl complex $[\text{K}\{\text{Al}(\text{NON})\}]_2$ in yields of approximately 75%, whereas the use of less forcing conditions generates the $\text{Al}\text{--}\text{Al}$ bonded dialane, $[\text{Al}(\text{NON})]_2$ (Fig. 1). Both compounds were characterized by spectroscopic and analytical techniques, and their solid-state structures determined by single crystal X-ray diffraction (Fig. 2, Extended Data Fig. 2 and Supplementary Information).

Given the well-known preference of aluminium for the +3 oxidation state¹, and the difficulty in locating hydrogen atoms by X-ray crystallography, it was important to rule out the presence of any metal-bound hydrogen atoms within the dimeric molecular unit of $[\text{K}\{\text{Al}(\text{NON})\}]_2$. In particular, given the presence of potassium counter-ions, we wished to rule out the formation of the dihydroaluminate $[\text{K}\{\text{H}_2\text{Al}(\text{NON})\}]_2$. Accordingly, $[\text{K}\{\text{H}_2\text{Al}(\text{NON})\}]_2$ was synthesized independently by the reaction of $(\text{NON})\text{AlI}$ with excess $\text{K}[\text{AlH}_4]$, and characterized by spectroscopic, analytical and crystallographic techniques (Extended Data Fig. 3). These measurements revealed considerable differences from the corresponding data of $[\text{K}\{\text{Al}(\text{NON})\}]_2$. These included shifted ^1H NMR resonances for the diamido-dimethylxanthene ligand backbone, and an additional broad signal at $\delta_{\text{H}} = 3.88$ p.p.m. associated with the aluminium-bound hydrogen atoms in $[\text{K}\{\text{H}_2\text{Al}(\text{NON})\}]_2$; an additional $\text{Al}\text{--}\text{H}$ stretching band in the infrared spectrum at $1,688\text{ cm}^{-1}$ (Extended Data Fig. 4); and shorter $\text{Al}\text{--}\text{O}$ and $\text{Al}\text{--}\text{N}$ distances in the solid-state structure

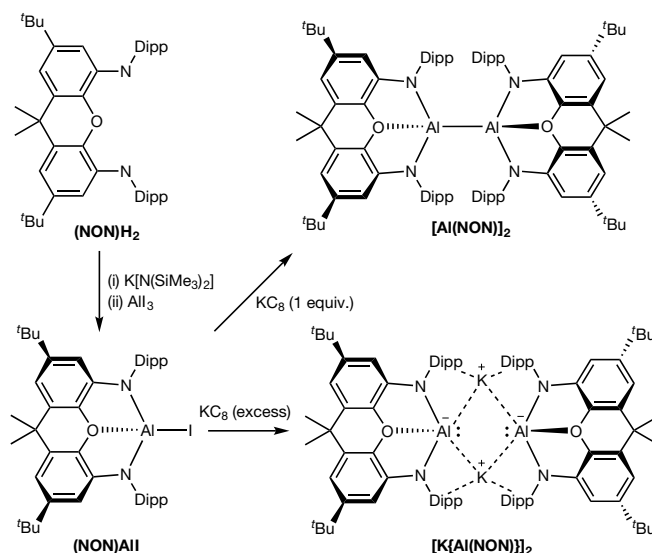


Fig. 1 | Syntheses of the potassium aluminyl compound $[\text{K}\{\text{Al}(\text{NON})\}]_2$ and the dialane $[\text{Al}(\text{NON})]_2$. Both compounds are prepared from $(\text{NON})\text{AlI}$ by reduction with potassium graphite.

¹Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford, UK. ²Department of Chemistry, NanoScience Center, University of Jyväskylä, Jyväskylä, Finland. *e-mail: simon.aldridge@chem.ox.ac.uk; jose.goicoechea@chem.ox.ac.uk

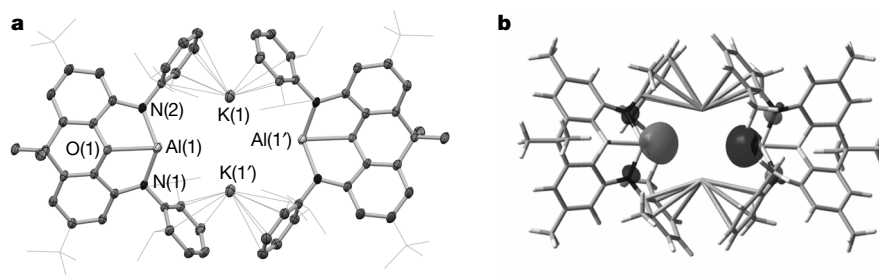


Fig. 2 | Geometric and electronic structure of $[K\{Al(NON)\}]_2$. **a**, Molecular structure of $[K\{Al(NON)\}]_2$ as determined by X-ray crystallography. Hydrogen atoms have been omitted and selected carbon atoms are shown in wireframe format for clarity; thermal ellipsoids have been drawn at the 35% probability level. Key distances (Å) and angles (°):

Al(1)...Al(1') 6.627(1), Al(1)...K(1) 4.070(1), Al(1)...K(1') 3.844(1), Al(1)–N(1) 1.963(2), Al(1)–N(2) 1.956(2), Al(1)–O(1) 2.279(2), N(1)–Al(1)–N(2) 128.1(1). **b**, The HOMO of $[K\{Al(NON')\}]_2$, as calculated by DFT. Isovalue = 0.05.

(mean values of 2.137 Å and 1.923 Å respectively, compared with 2.279(2) Å and 1.956(2) or 1.963(2) Å for $[K\{Al(NON)\}]_2$). Perhaps even more definitive is the additional finding that $[K\{Al(NON)\}]_2$ adds dihydrogen under ambient conditions in benzene solution (or at 2 atm pressure in the solid state), with its conversion into $[K\{H_2Al(NON)\}]_2$ being confirmed by in situ NMR measurements.

The structure of $[K\{Al(NON)\}]_2$ in the solid state is shown by crystallographic studies to be a centrosymmetric dimer, featuring two formally anionic $[Al(NON)]^-$ units held together through flanking potassium–arene contacts involving the two potassium counter-ions ($d(K...C) = 3.226(3)–3.474(3)$ Å). The non-bonded Al...Al separation is more than double the length of the formal Al–Al single bond found in $[Al(NON)]_2$ (6.627(1) compared with 2.646(1) Å)²⁰, whereas the K...Al contacts (3.844(1) and 4.070(1) Å) are comparable to (or slightly longer than) those reported, for example, for the tetragallium cluster $K_2[Ar_2Ga_4]$, which also features potassium counter-ions sandwiched between flanking aryl rings (3.471(1)–3.833(1) Å; Ar = $C_6H_3(C_6H_2)^{iPr}_{3-2,4,6}$)₂²¹.

The aluminium centre in $[K\{Al(NON)\}]_2$ features a flattened pyramidal coordination sphere ($\angle(N-Al-N) = 128.1(1)^\circ$; $\angle(N-Al-O) = 72.9(1), 72.5(1)^\circ$). The Al–N distances (1.956(2) and 1.963(2) Å) are consistent with an Al(I) compound: the corresponding bond lengths measured for $[Al(NON)]_2$ and $(NON)AlI$ are successively shorter (1.895(2)–1.901(2) Å and 1.846(2) Å, respectively), in line with the reduced covalent radii of Al(II) and Al(III) (compared with that of Al(I))^{1,22}. In addition, whereas the Al–O distances associated with the neutral ether donor are very similar for $[Al(NON)]_2$ and $(NON)AlI$ (1.976(2) or 1.981(2) Å and 1.967(2) Å, respectively), that measured

for $[K\{Al(NON)\}]_2$ is markedly longer (2.279(2) Å), consistent with the reduced Lewis acidity expected for a formally anionic metal centre. This metal–oxygen distance is, however, noticeably shorter than that measured for the isostructural gallium analogue $[K\{Ga(NON)\}]_2$ (2.542(2) Å; Extended Data Fig. 5) and the associated ‘puckering’ of the dimethylxanthene backbone is more pronounced (the angle between least squares planes of aromatic rings is 38.6° , compared with 25.9° for $[K\{Ga(NON)\}]_2$). Both of these observations suggest that the dative oxygen \rightarrow metal interaction is structurally more important for the aluminium system, with the implied population of the Al-centred p_z orbital potentially leading to a relatively wide HOMO–LUMO gap (the energy difference between the highest occupied and lowest unoccupied molecular orbitals).

The electronic structure of $[K\{Al(NON)\}]_2$ was probed using density functional theory (DFT) calculations, both on the monomeric alumanyl anion $[Al(NON)]^-$ and on the model dimeric system $[K\{Al(NON')\}]_2$ (in which the iPr and tBu substituents were replaced by Me for computational efficiency). These calculations suggest that the HOMO–LUMO gap for the $[Al(NON)]^-$ fragment is approximately 338 kJ mol^{-1} (3.50 eV), and that a singlet electronic ground state is therefore predicted ($\Delta E_{st} = 166\text{ kJ mol}^{-1}$, 1.72 eV)^{17,18}. Presumably, a major contributory factor here is the fact that the LUMO + 3 (rather than the LUMO) features the primary contribution from the aluminium p_z orbital, and that this orbital is effectively Al–O σ^* antibonding in character (consistent with the observed Al–O distance).

The energy of the HOMO calculated for the (hypothetical) monomeric $[Al(NON)]^-$ anion in the gas phase is very high at -101 kJ mol^{-1} (-1.05 eV), but this value decreases to -395 kJ mol^{-1} (-4.10 eV) for

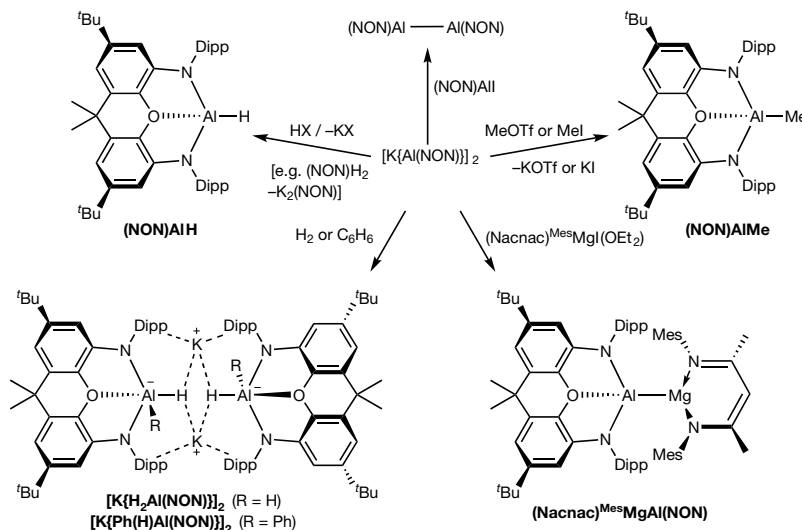


Fig. 3 | Exploitation of the aluminium-centred nucleophilic reactivity of $[K\{Al(NON)\}]_2$. The formation of Al–H, Al–C and Al–M bonds

is achieved by the reaction of the potassium alumanyl reagent with appropriate electrophiles.

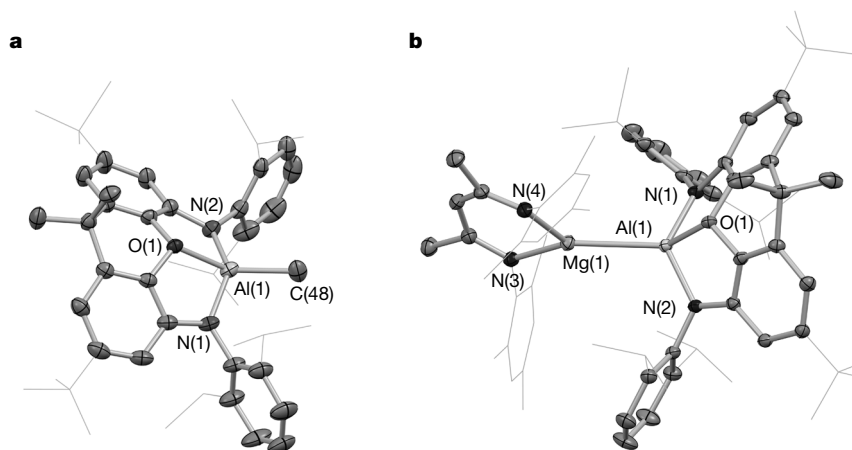


Fig. 4 | Molecular structures of an aluminium alkyl and a magnesium aluminyl compound formed via reactions of $[K\{Al(NON)\}]_2$, as determined by X-ray crystallography. Hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity; thermal ellipsoids have been drawn at the 35% probability level. **a**, (NON)AlMe. Key bond lengths (Å) and angles (°): Al(1)–C(48) 2.029(3),

Al(1)–N(1) 1.870(2), Al(1)–N(2) 1.867(2), Al(1)–O(1) 1.994(2), N(1)–Al(1)–N(2) 138.3(1). **b**, (Nacnac)^{Mes}MgAl(NON). Key bond lengths (Å) and angles (°): Al(1)–Mg(1) 2.696(1), Al(1)–N(1) 1.918(2), Al(1)–N(2) 1.904(1), Al(1)–O(1) 1.992(1), Mg(1)–N(3) 2.033(2), Mg(1)–N(4) 2.035(1), N(1)–Al(1)–N(2) 126.4(1).

the $[K\{Al(NON')\}]_2$ dimer—the HOMO in this case being the out-of-phase combination of Al-centred lone pairs (Fig. 2). The idea that the dimeric K^+ -bridged structure offers considerably enhanced stability gains additional support from diffusion-ordered spectroscopy (DOSY) NMR measurements carried out in C_6D_6 solution. The hydrodynamic radius determined for $[K\{Al(NON)\}]_2$ (9.7 Å) can be compared to values of 7.7 Å and 8.9 Å measured for monomeric (NON)AlI and dinuclear $[Al(NON)]_2$, suggesting that the dimeric structure of the potassium aluminyl compound is retained in the solution phase.

Nonetheless, the HOMO energy for $[K\{Al(NON')\}]_2$ is markedly higher than that calculated for the charge-neutral Al(I) compound Al(Nacnac)^{Dipp} using the same method (-453 kJ mol^{-1} , -4.70 eV)⁹. Taken together with the slightly higher contribution of the aluminium 3p orbital to the lone pair (24% compared with 10% for Al(Nacnac)^{Dipp}), these data suggest that $[K\{Al(NON)\}]_2$ should have greater potential to act as an aluminium-centred nucleophile. This assertion can be verified experimentally by the use of $[K\{Al(NON)\}]_2$ in a range of unprecedented aluminium–element bond-forming reactions with electrophilic partners (Fig. 3). Its reaction with methyl triflate or methyl iodide to give (NON)AlMe (Fig. 4) demonstrates a new approach for the formation of aluminium alkyls that complements conventional methods using hydroalumination chemistry or aluminium electrophiles¹. In a similar manner, reactions with Brønsted acids can be used to generate the monomeric hydride (NON)AlH via protonation at aluminium (Extended Data Fig. 6). Covalent metal–aluminium bonds can be constructed in a similar manner. The combination of $[K\{Al(NON)\}]_2$ with (NON)AlI defines an alternative strategy for the formation of the dialuminium compound (NON)Al–Al(NON), whereas its reaction with (Nacnac)^{Mes}MgI(OEt₂) (where (Nacnac)^{Mes} = (NMesCMe)₂CH, and Mes = 2,4,6-Me₃C₆H₂) results in the formation of the magnesium aluminyl compound (Nacnac)^{Mes}MgAl(NON) (Figs. 3 and 4). The latter features an unsupported Mg–Al bond, with the associated bond length (2.696(1) Å) comparable to the sum of the respective covalent radii (2.62 Å)^{22,23}. The metal–metal distance is also in line with that reported for the single Mg–Mg covalent bond in the Mg(I) dimer $[Mg(Nacnac)^{Dipp}]_2$ ($d(\text{Mg–Mg}) = 2.851(1) \text{ Å}$)²⁴, with due allowance for the differing radii of aluminium and magnesium ($\Delta r_{\text{cov}} = 0.2 \text{ Å}$)²². As such, whereas $[K\{Al(NON)\}]_2$ represents an aluminyl analogue of organo-group 1 nucleophiles of the type RLi and RK⁴, (Nacnac)^{Mes}MgAl(NON) can be thought of as the corresponding counterpart of a Grignard reagent, RMgX¹⁹.

Despite having no precedent as a structurally authenticated aluminyl species, $[K\{Al(NON)\}]_2$ is stable for several days at 300 K, both in

benzene solution and in the solid state. At 330 K, however, clean conversion to a single species $[K\{\text{Ph(H)Al(NON)}\}]_2$ is observed, via formal oxidative cleavage of a C–H bond of benzene at Al(I) (Extended Data Fig. 7). To our knowledge this represents a first demonstration of the intermolecular oxidative addition of a C–H bond in benzene at a single well-defined main-group metal centre. Although main-group compounds that activate benzene by formal deprotonation are known^{25,26}, oxidative cleavage is preceded only for more reactive C–H bonds¹⁴. In summary, we have synthesized and structurally characterized the first example, to our knowledge, of an anionic Al(I) (aluminyl) system—a compound that offers potential as a nucleophilic reagent in the formation of unsupported aluminium–element bonds and in the activation of typically inert small-molecule substrates.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0037-y>.

Received: 6 November 2017; Accepted: 16 February 2018;

Published online 16 April 2018.

1. Aldridge, S. & Downs, A. J. (eds) *The Group 13 Metals Aluminium, Gallium, Indium and Thallium: Chemical Patterns and Peculiarities* (Wiley, Chichester, 2011).
2. Helmboldt, O. et al. *Ullmann's Encyclopedia of Industrial Chemistry: Aluminum Compounds, Inorganic* (Wiley VCH, Weinheim, 2007).
3. Lappert, M., Protchenko, A., Power, P. & Seeber, A. *Metal Amide Chemistry* (Wiley, Chichester, 2009).
4. Rappoport, Z. & Marek, I. (eds) *The Chemistry of Organolithium Compounds* (Wiley–Blackwell, Chichester, 2004).
5. Segawa, Y., Yamashita, M. & Nozaki, K. Boryllithium: isolation, characterization, and reactivity as a boryl anion. *Science* **314**, 113–115 (2006).
6. Dohmeier, C., Loos, D. & Schnöckel, H. Aluminum(I) and gallium(I) compounds: syntheses, structures, and reactions. *Angew. Chem. Int. Ed.* **35**, 129–149 (1996).
7. Nagendran, S. & Roesky, H. The chemistry of aluminum(I), silicon(II) and germanium(II). *Organometallics* **27**, 457–492 (2008).
8. Jones, C. & Stasch, A. in *The Group 13 Metals Aluminium, Gallium, Indium and Thallium: Chemical Patterns and Peculiarities* (eds Aldridge, S. & Downs, A. J.) 285–341 (Wiley, Chichester, 2011).
9. Cui, C. et al. Synthesis and structure of a monomeric aluminum(I) compound $[(\text{HC}(\text{CMe}(\text{NAr})_2)\text{Al}]$ (Ar = 2,6-*i*-Pr₂C₆H₃): a stable aluminum analogue of a carbene. *Angew. Chem. Int. Ed.* **39**, 4274–4276 (2000).
10. Li, X., Cheng, X., Song, H. & Cui, C. Synthesis of $\text{HC}[(\text{CMe}(\text{NAr})_2)]_2\text{Al}$ (Ar = 2,6-*i*-Pr₂C₆H₃) and its reaction with isocyanides, a bulky azide, and H₂O. *Organometallics* **26**, 1039–1043 (2007).
11. Linti, G. & Schnöckel, H. Low valent aluminium and gallium compounds — structural variety and coordination modes to transition metal fragments. *Coord. Chem. Rev.* **206–207**, 285–319 (2000).

12. Asay, M., Jones, C. & Driess, M. *N*-heterocyclic carbene analogues with low-valent group 13 and group 14 elements: syntheses, structures, and reactivities of a new generation of multitailented ligands. *Chem. Rev.* **111**, 354–396 (2011).
13. González-Gallardo, S., Bollermann, T., Fischer, R. A. & Murugavel, R. Cyclopentadiene based low-valent group 13 metal compounds: ligands in coordination chemistry and link between metal rich molecules and intermetallic materials. *Chem. Rev.* **112**, 3136–3170 (2012).
14. Chu, T., Korobkov, I. & Nikonov, G. I. Oxidative addition of σ bonds to an Al(I) center. *J. Am. Chem. Soc.* **136**, 9195–9202 (2014).
15. Crimmin, M. R., Butler, M. J. & White, A. J. P. Oxidative addition of carbon–fluorine and carbon–oxygen bonds to Al(I). *Chem. Commun.* **51**, 15994–15996 (2015).
16. Chu, T., Boyko, Y., Korobkov, I. & Nikonov, G. I. Transition metal-like oxidative addition of C–F and C–O bonds to an aluminum(I) center. *Organometallics* **34**, 5363–5365 (2015).
17. Sundermann, A., Reiher, M. & Schoeller, W. W. Isoelectronic Arduengo-type carbene analogues with the group IIIa elements boron, aluminum, gallium, and indium. *Eur. J. Inorg. Chem.* **1998**, 305–310 (1998).
18. Tuononen, H. M., Roesler, R., Dutton, J. L. & Ragogna, P. J. Electronic structures of main-group carbene analogues. *Inorg. Chem.* **46**, 10693–10706 (2007).
19. Westrum, L. J. & Rakita, P. E. (eds) *Handbook of Grignard Reagents* 2nd edn (CRC Press, Boca Raton, 2015).
20. Uhl, W. Organoelement compounds possessing Al–Al, Ga–Ga, In–In, and Tl–Tl single bonds. *Adv. Organomet. Chem.* **51**, 53–108 (2004).
21. Twamley, B. & Power, P. P. Synthesis of the square-planar gallium species $K_2[Ga_4(C_6H_3-2,6-Trip_2)_2]$ (Trip = $C_6H_2-2,4,6-Pr_3$): the role of aryl–alkali metal ion interactions in the structure of gallium clusters. *Angew. Chem. Int. Ed.* **39**, 3500–3503 (2000).
22. Cordero, B. et al. Covalent radii revisited. *Dalton Trans.* **21**, 2832–2838 (2008).
23. Bakewell, C., Ward, B. J., White, A. J. P. & Crimmin, M. R. A combined experimental and computational study on the reaction of fluoroarenes with Mg–Mg, Mg–Zn, Mg–Al and Al–Zn bonds. *Chem. Sci.* **9**, 2348–2356 (2018).
24. Green, S. P., Jones, C. & Stasch, A. Stable magnesium(I) compounds with Mg–Mg bonds. *Science* **318**, 1754–1757 (2007).
25. Martínez-Martínez, A. J., Kennedy, A. R., Mulvey, R. E. & O'Hara, C. T. Directed ortho-meta'- and meta-meta'-dimetalations: a template base approach to deprotonation. *Science* **346**, 834–837 (2014).
26. Ohsato, T. et al. A potassium diboryllithate: synthesis, bonding properties, and the deprotonation of benzene. *Angew. Chem. Int. Ed.* **55**, 11426–11430 (2016).

Acknowledgements This work was supported by the SCG-Oxford Centre of Excellence. P.V. thanks the Magnus Ehrnrooth and Emil Aaltonen Foundations for postdoctoral funding. We thank the University of Oxford Advanced Research Computing facility, and N. Rees and H. Tuononen for assistance with NMR and quantum chemical studies, respectively.

Author contributions J.H. carried out the synthetic and reaction studies, P.V. carried out the computational analyses, J.H. and J.M.G. conducted the crystallographic studies, and J.M.G. and S.A. wrote the manuscript and managed the project.

Competing interests: The authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41586-018-0037-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0037-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to S.A. or J.M.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

General considerations. All manipulations were carried out using standard Schlenk-line or dry-box techniques under an atmosphere of argon or dinitrogen. Solvents were degassed by sparging with argon and dried by passing through a column of the appropriate drying agent. NMR spectra were measured in benzene- d_6 (which was dried over potassium) or THF- d_8 (which was dried over LiAlH_4), with the solvent then being distilled under reduced pressure and stored under argon in Teflon valve ampoules. NMR samples were prepared under argon in 5-mm Wilmad 507-PP tubes fitted with J. Young Teflon valves. ^1H and $^{13}\text{C}\{^1\text{H}\}$ NMR spectra were recorded on a Bruker Avance III HD NanoBay 400 MHz or Bruker Avance III 500 MHz spectrometer at ambient temperature, referenced internally to residual proto-solvent (^1H) or solvent (^{13}C) resonances and are reported relative to tetramethylsilane ($\delta=0$ p.p.m.). Assignments were confirmed using two-dimensional ^1H - ^1H and ^{13}C - ^1H NMR correlation experiments. Chemical shifts are quoted in δ (p.p.m.) and coupling constants (J) in Hz. Elemental analyses were carried out by London Metropolitan University.

Starting materials. $(\text{NON})\text{H}_2$ ²⁷, $(\text{Nacnac})^{\text{Mes}}\text{MgI}(\text{OEt})_2$ ²⁸ and KAlH_4 ²⁹ were prepared according to literature methods. AlI_3 and GaI_3 were prepared in situ by sonicating a mixture of the appropriate metal with 1.5 equivalents of I_2 in dry toluene under an atmosphere of argon until the solution became colourless. All other reagents were used as received.

Syntheses of novel compounds. *(NON)AlI*. To a solution of $\text{K}_2(\text{NON})$ (2.00 g, 2.67 mmol) in toluene (30 ml) was added a solution of AlI_3 (1.09 g, 2.67 mmol) in toluene (30 ml) at -78°C for 15 min. The reaction mixture was slowly warmed to room temperature and stirred for a further 16 h. The reaction mixture was filtered and volatiles from the filtrate were removed in vacuo to give an off-white solid. This solid was washed with hexane (2×20 ml) to give $(\text{NON})\text{AlI}$ as a white powder (1.55 g, 70% yield). X-ray quality crystals of $(\text{NON})\text{AlI}$ were obtained by recrystallizing this solid from warm toluene. ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = 1.11$ (d, $^3J_{\text{HH}} = 5.8$ Hz, 12 H, $\text{CH}(\text{CH}_3)_2$), 1.15 (s, 18 H, $\text{C}(\text{CH}_3)_3$), 1.40 (d, $^3J_{\text{HH}} = 6.0$ Hz, 12 H, $\text{DippCH}(\text{CH}_3)_2$), 1.51 (s, 6 H, $\text{XA-C}(\text{CH}_3)_2$), 3.54 (br, 4 H, $\text{CH}(\text{CH}_3)_2$), 6.39 (s, 2 H, XA-o-CH), 6.78 (s, 2 H, XA-p-CH), 7.27 (s, 6 H, ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (125.8 MHz, C_6D_6): $\delta = 24.7$, 25.6 ($\text{CH}(\text{CH}_3)_2$), 27.4 ($\text{XA-C}(\text{CH}_3)_2$), 29.6 ($\text{CH}(\text{CH}_3)_2$), 31.6 ($\text{C}(\text{CH}_3)_3$), 35.2 ($\text{C}(\text{CH}_3)_3$), 37.4 ($\text{XA-C}(\text{CH}_3)_2$), 109.1, 112.6, 124.7, 127.3, 128.0, 128.4, 133.2, 139.2, 140.8, 142.6, 147.0, 149.9 (Ar-C); anal. calc. for $\text{C}_{47}\text{H}_{62}\text{AlIN}_2\text{O}$: C 68.43%, H 7.58%, N 3.40%, found: C 68.36%, H 7.68%, N 3.36%.

[Al(NON)]₂. Method A: to a suspension of KC_8 (0.049 g, 0.364 mmol) in toluene (25 ml) was added a solution of $(\text{NON})\text{AlI}$ (0.300 g, 0.364 mmol) in toluene (25 ml) at room temperature. The reaction mixture was stirred for a further 16 h at room temperature whereupon volatiles were removed in vacuo. The residue was extracted with pentane (15 ml), the extract filtered and volatiles removed in vacuo to give $[\text{Al}(\text{NON})]_2$ as an off-white solid (0.120 g, 47%). X-ray quality crystals of $[\text{Al}(\text{NON})]_2$ were obtained by crystallizing this solid from hexane. Method B: to a solution of $[\text{K}[\text{Al}(\text{NON})]]_2$ (0.200 g, 0.136 mmol) in toluene (10 ml) was added a solution of $(\text{NON})\text{AlI}$ (0.246 g, 0.298 mmol) in toluene (10 ml) at room temperature. The reaction mixture was stirred for a further 16 h at room temperature whereupon volatiles were removed in vacuo. The residue was extracted with pentane (15 ml), the extract filtered and volatiles removed in vacuo to give $[\text{Al}(\text{NON})]_2$ as an off-white solid (0.325 g, 86%). ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = -0.12$ (d, $^3J_{\text{HH}} = 6.8$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 0.11 (d, $^3J_{\text{HH}} = 6.9$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 0.66 (d, $^3J_{\text{HH}} = 6.7$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 0.75 (d, $^3J_{\text{HH}} = 6.3$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 1.09 (d, $^3J_{\text{HH}} = 6.9$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 1.15 (s, 36 H, $\text{C}(\text{CH}_3)_3$), 1.16 (d, $^3J_{\text{HH}} = 6.9$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 1.35 (d, $^3J_{\text{HH}} = 6.5$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 1.58 (d, $^3J_{\text{HH}} = 7.1$ Hz, 6 H, $\text{CH}(\text{CH}_3)_2$), 1.65 (s, 6 H, $\text{XA-C}(\text{CH}_3)_2$), 1.81 (s, 6 H, $\text{XA-C}(\text{CH}_3)_2$), 2.01 (sept., $^3J_{\text{HH}} = 6.7$ Hz, 2 H, $\text{CH}(\text{CH}_3)_2$), 3.02 (sept., $^3J_{\text{HH}} = 6.8$ Hz, 2 H, $\text{CH}(\text{CH}_3)_2$), 3.11 (sept., $^3J_{\text{HH}} = 6.7$ Hz, 2 H, $\text{CH}(\text{CH}_3)_2$), 3.90 (sept., $^3J_{\text{HH}} = 6.6$ Hz, 2 H, $\text{CH}(\text{CH}_3)_2$), 6.17 (s, 2 H, XA-o-CH), 6.57 (s, 2 H, XA-o-CH), 6.63 (s, 2 H, XA-p-CH), 6.68 (s, 2 H, XA-p-CH), 6.92–7.32 (m, 12 H ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, C_6D_6): $\delta = 22.2$ ($\text{CH}(\text{CH}_3)_2$), 23.2 ($\text{XA-C}(\text{CH}_3)_2$), 23.6, 23.8, 24.5, 24.9, 25.0, 26.9 ($\text{CH}(\text{CH}_3)_2$), 27.2, 29.0 ($\text{CH}(\text{CH}_3)_2$), 29.5 ($\text{XA-C}(\text{CH}_3)_2$), 29.9 ($\text{CH}(\text{CH}_3)_2$), 30.5 ($\text{CH}(\text{CH}_3)_2$), 30.6 ($\text{CH}(\text{CH}_3)_2$), 31.6, 31.7 ($\text{C}(\text{CH}_3)_3$), 35.0, 35.1 ($\text{C}(\text{CH}_3)_3$), 39.6 ($\text{XA-C}(\text{CH}_3)_2$), 108.0, 109.5, 112.7, 114.2, 123.3, 124.1, 124.3, 126.0, 126.1, 126.6, 137.0, 139.2, 141.2, 142.1, 142.9, 144.5, 145.4, 145.8, 147.2, 147.5, 148.0, 148.3, 148.7, 150.1 (Ar-C); anal. calc. for $\text{C}_{94}\text{H}_{124}\text{Al}_2\text{N}_4\text{O}_2$: C 80.88%, H 8.95%, N 4.01%, found: C 80.66%, H 9.06%, N 4.19%.

[K[Al(NON)]]₂. To a suspension of KC_8 (0.410 g, 3.03 mmol) in toluene (25 ml) was added a solution of $(\text{NON})\text{AlI}$ (1.00 g, 1.21 mmol) in toluene (25 ml) at room temperature. The reaction mixture was stirred for 16 h at room temperature producing a colour change from colourless to yellow. The reaction mixture was filtered and volatiles from the filtrate were removed in vacuo to give $[\text{K}[\text{Al}(\text{NON})]]_2$ as a bright yellow powder (0.670 g, 76% yield). X-ray quality crystals of $[\text{K}[\text{Al}(\text{NON})]]_2$ were obtained by recrystallizing this bright yellow powder from warm benzene. ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = 1.11$ (d, $^3J_{\text{HH}} = 6.6$ Hz, 24 H, $\text{CH}(\text{CH}_3)_2$),

1.16 (d, $^3J_{\text{HH}} = 6.6$ Hz, 24 H, $\text{CH}(\text{CH}_3)_2$), 1.27 (s, 36 H, $\text{C}(\text{CH}_3)_3$), 1.69 (s, 12 H, $\text{XA-C}(\text{CH}_3)_2$), 3.70 (sept., $^3J_{\text{HH}} = 6.6$ Hz, 8 H, $\text{CH}(\text{CH}_3)_2$), 6.15 (d, $^4J_{\text{HH}} = 1.6$ Hz, 4 H, XA-o-CH), 6.76 (s, $^4J_{\text{HH}} = 1.6$ Hz, 4 H, XA-p-CH), 6.99–7.14 (m, 12 H, ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, C_6D_6): $\delta = 24.9$, 25.1 ($\text{CH}(\text{CH}_3)_2$), 28.1 ($\text{CH}(\text{CH}_3)_2$), 28.9 ($\text{XA-C}(\text{CH}_3)_2$), 32.1 ($\text{C}(\text{CH}_3)_3$), 35.1 ($\text{C}(\text{CH}_3)_3$), 37.1 ($\text{XA-C}(\text{CH}_3)_2$), 106.2, 109.2, 124.1, 125.3, 128.6, 133.2, 142.4, 142.8, 146.8, 147.6, 149.5 (Ar-C); IR ν (cm^{-1} , Nujol): 1,642(m), 1,584(m), 1,485(m), 1,442(s), 1,419(s), 1,362(m), 1,321(s), 1,255(m), 1,200(m), 1,175(s), 1,135(m), 1,112(m), 1,102(m), 1,055(m), 1,044(m), 1,027(m), 1,014(s), 990(m), 943(s), 934(s), 909(m), 863(m), 848(m), 801(m), 794(s), 773(s), 766(m), 729(m), 675(m), 655(m), 637(m), 620(m), 582(s), 569(s); anal. calc. for $\text{C}_{94}\text{H}_{124}\text{Al}_2\text{K}_2\text{N}_4\text{O}_2$: C 76.59%, H 8.48%, N 3.80%, found: C 76.74%, H 8.32%, N 3.59%.

(NON)AlH. To a solution of $[\text{K}[\text{Al}(\text{NON})]]_2$ (0.200 g, 0.136 mmol) in toluene (10 ml) was added a solution of $(\text{NON})\text{H}_2$ (0.092 g, 0.136 mmol) dropwise at -78°C . The solution was slowly warmed to room temperature and stirred for a further 2 h, whereupon volatiles were removed in vacuo. The residue was extracted with ice-cold hexane (15 ml), the extract filtered and volatiles removed from the filtrate in vacuo to give $(\text{NON})\text{AlH}$ as a colourless solid (0.167 g, 88%). X-ray quality crystals of $(\text{NON})\text{AlH}$ were obtained by crystallizing this solid from warm toluene. ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = 1.16$ (d, $^3J_{\text{HH}} = 6.8$ Hz, 12 H, $\text{CH}(\text{CH}_3)_2$), 1.19 (s, 18 H, $\text{C}(\text{CH}_3)_3$), 1.25 (d, $^3J_{\text{HH}} = 6.8$ Hz, 12 H, $\text{CH}(\text{CH}_3)_2$), 1.52 (s, 6 H, $\text{C}(\text{CH}_3)_2$), 3.56 (sept., $^3J_{\text{HH}} = 6.8$ Hz, 4 H, $\text{CH}(\text{CH}_3)_2$), 4.99 (br, 1 H, AlH), 6.43 (d, $^4J_{\text{HH}} = 1.9$ Hz, 2 H, XA-o-CH), 6.76 (d, $^4J_{\text{HH}} = 1.9$ Hz, 2 H, XA-p-CH), 7.27 (s, 6 H ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, C_6D_6): $\delta = 24.9$, 25.6 ($\text{CH}(\text{CH}_3)_2$), 27.5 ($\text{XA-C}(\text{CH}_3)_2$), 29.0 ($\text{CH}(\text{CH}_3)_2$), 31.8 ($\text{C}(\text{CH}_3)_3$), 35.3 ($\text{C}(\text{CH}_3)_3$), 37.6 ($\text{XA-C}(\text{CH}_3)_2$), 108.1, 110.6, 124.6, 127.0, 128.4, 128.6, 133.4, 138.8, 141.1, 142.9, 147.6, 150.0 (Ar-C); IR ν (cm^{-1} , Nujol): 1,875(s, Al-H), 1,640(s), 1,583(m), 1,418(m), 1,364(s), 1,323(m), 1,310(m), 1,254(s), 1,200(m), 1,175(s), 1,100(m), 1,015(m), 935(m), 796(m), 727(m), 714(s); anal. calc. for $\text{C}_{47}\text{H}_{63}\text{AlN}_2\text{O}$: C 80.76%, H 9.08%, N 3.86%, found: C 80.72%, H 9.16%, N 3.94%.

(NON)AlMe. To a solution of $[\text{K}[\text{Al}(\text{NON})]]_2$ (0.200 g, 0.136 mmol) in toluene (10 ml) was added a solution of MeI (0.046 g, 0.326 mmol) in toluene (10 ml) at room temperature. The reaction instantly changed colour from bright yellow to colourless upon addition. The mixture was stirred for a further 2 h at room temperature whereupon volatiles were removed in vacuo. The residue was extracted with toluene (15 ml), the extract filtered and volatiles removed from the filtrate in vacuo to give $(\text{NON})\text{AlMe}$ as an off-white solid (0.184 g, 95%). X-ray quality crystals of $(\text{NON})\text{AlMe}$ were obtained by recrystallizing this solid from warm toluene. ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = -0.35$ (s, 3 H, AlCH_3), 1.12 (d, $^3J_{\text{HH}} = 6.8$ Hz, 12 H, $\text{CH}(\text{CH}_3)_2$), 1.18 (s, 18 H, $\text{C}(\text{CH}_3)_3$), 1.26 (d, $^3J_{\text{HH}} = 6.8$ Hz, 12 H, $\text{CH}(\text{CH}_3)_2$), 1.58 (s, 6 H, $\text{XA-C}(\text{CH}_3)_2$), 3.53 (sept., $^3J_{\text{HH}} = 6.8$ Hz, 4 H, $\text{CH}(\text{CH}_3)_2$), 6.34 (d, $^4J_{\text{HH}} = 1.8$ Hz, 2 H, XA-o-CH), 6.76 (s, $^4J_{\text{HH}} = 1.8$ Hz, 2 H, XA-p-CH), 7.24 (s, 6 H, ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, C_6D_6): $\delta = -12.9$ (AlCH_3), 24.5, 25.8 ($\text{CH}(\text{CH}_3)_2$), 27.4 ($\text{XA-C}(\text{CH}_3)_2$), 29.1 ($\text{CH}(\text{CH}_3)_2$), 31.7 ($\text{C}(\text{CH}_3)_3$), 35.2 ($\text{C}(\text{CH}_3)_3$), 37.6 ($\text{XA-C}(\text{CH}_3)_2$), 107.9, 111.5, 124.6, 126.7, 133.8, 140.7, 141.7, 143.5, 147.1, 149.5 (Ar-C); anal. calc. for $\text{C}_{48}\text{H}_{65}\text{AlN}_2\text{O}$: C 80.85%, H 9.19%, N 3.93%, found: C 80.97%, H 9.28%, N 3.78%.

(Nacnac)^{Mes}MgAl(NON). To a solution of $[\text{K}[\text{Al}(\text{NON})]]_2$ (0.200 g, 0.137 mmol) in benzene (15 ml) was added a solution of $[(^{\text{Mes}}\text{Nacnac})\text{MgI}(\text{OEt})_2]$ (0.167 g, 0.299 mmol) in benzene (15 ml) at room temperature over 5 min. The reaction mixture was stirred for 16 h at room temperature, producing a colour change from yellow to colourless. Volatiles were removed in vacuo, the residue was extracted with toluene (20 ml), the extract filtered, concentrated in vacuo (to around 5 ml) and slowly cooled to 5°C overnight to give $(\text{Nacnac})^{\text{Mes}}\text{MgAl}(\text{NON})$ as colourless crystals (0.245 g, 86% yield). ^1H NMR (400 MHz, C_6D_6 , 298 K): $\delta = 1.00$ (d, $^3J_{\text{HH}} = 6.3$ Hz, 6 H, $\text{DippCH}(\text{CH}_3)_2$), 1.10–1.12 (m, 18 H, $\text{DippCH}(\text{CH}_3)_2$), 1.16 (s, 18 H, $\text{C}(\text{CH}_3)_3$), 1.39 (s, 3 H, NCCH_3), 1.58 (s, 6 H, o-CH_3), 1.60 (s, 3 H, NCCH_3), 1.71 (s, 3 H, XA-CH_3), 1.76 (s, 3 H, XA-CH_3), 2.14 (s, 3 H, p-CH_3), 2.19 (s, 3 H, p-CH_3), 2.32 (s, 6 H, o-CH_3), 3.35 (sept., $^3J_{\text{HH}} = 6.3$ Hz, 2 H, $\text{DippCH}(\text{CH}_3)_2$), 3.57 (sept., $^3J_{\text{HH}} = 6.4$ Hz, 2 H, $\text{DippCH}(\text{CH}_3)_2$), 4.86 (s, 1 H, NCCH_3), 6.19 (s, 2 H, XA-o-CH), 6.58 (s, 2 H, Mes-m-CH), 6.69 (s, 2 H, XA-p-CH), 6.84 (s, 2 H, Mes-m-CH), 7.13–7.18 (m, 6 H, ArH); $^{13}\text{C}\{^1\text{H}\}$ NMR (125.8 MHz, C_6D_6): $\delta = 18.4$, 19.4 (Mes-o-CH_3), 20.9, 21.0 (Mes-p-CH_3), 22.5 (XA-CH_3), 22.7, 23.9 (NCCH_3), 24.0, 25.1, 26.2, 26.6 ($\text{DippCH}(\text{CH}_3)_2$), 28.7, 29.3 ($\text{DippCH}(\text{CH}_3)_2$), 31.3 (XA-CH_3), 31.8 ($\text{C}(\text{CH}_3)_3$), 35.1 ($\text{C}(\text{CH}_3)_3$), 37.9 ($\text{XA-C}(\text{CH}_3)_2$), 96.6 (NCCH_3), 106.5, 110.8, 123.7, 124.6, 125.9, 127.9, 128.2, 129.5, 129.7, 130.8, 132.2, 133.1, 133.5, 134.5, 141.3, 143.4, 143.5, 143.6, 145.6, 146.5, 148.3, 148.7 (Ar-C), 168.1, 168.6 (NCCH_3); anal. calc. for $\text{C}_{70}\text{H}_{91}\text{AlMgN}_4\text{O}$: C 79.63%, H 8.69%, N 5.31%, found: C 79.64%, H 8.56%, N 5.14%.

[K(H₂Al(NON))]₂. Method A: to a suspension of $\text{K}[\text{AlH}_4]$ (0.127 g, 1.82 mmol) in benzene (10 ml) was added a solution of $(\text{NON})\text{AlI}$ (0.300 g, 0.364 mmol) in benzene at room temperature. The reaction mixture was heated to reflux and stirred for a further 48 h. After allowing the reaction mixture to cool to room temperature, it was filtered and volatiles were removed from the filtrate in

vacuo to give $[K\{H_2Al(NON)\}]_2$ as a colourless solid (0.258 g, 96%). Method B: a solution of $[K\{Al(NON)\}]_2$ (0.200 g, 0.136 mmol) in toluene (10 ml) was prepared in a 50 ml J. Young sample flask. The solution was frozen, the atmosphere evacuated (5.0×10^{-3} mbar) and refilled with H_2 (2.0 bar). The flask was slowly warmed to room temperature where it was left to stand for 5 days, producing a colour change from yellow to colourless. Volatiles were removed in vacuo to give $[K\{H_2Al(NON)\}]_2$ as a colourless solid (0.198 g, 99%). X-ray quality crystals were obtained by recrystallizing this solid from warm benzene. 1H NMR (400 MHz, C_6D_6 , 298 K): δ = 1.10 (d, $^3J_{HH}$ = 6.6 Hz, 24 H, $CH(CH_3)_2$), 1.15 (d, $^3J_{HH}$ = 6.6 Hz, 24 H, $CH(CH_3)_2$), 1.26 (s, 36 H, $C(CH_3)_3$), 1.60 (s, 12 H, $C(CH_3)_2$), 3.73 (sept., $^3J_{HH}$ = 6.6 Hz, 8 H, $CH(CH_3)_2$), 3.88 (br, 4 H, AlH_2), 6.07 (s, 4 H, $XA-o-CH$), 6.71 (s, 4 H, $XA-p-CH$), 7.01–7.10 (m, 12 H ArH); $^{13}C\{^1H\}$ NMR (126 MHz, C_6D_6): δ = 25.1, 25.6 ($CH(CH_3)_2$), 28.1 ($XA-C(CH_3)_2$), 28.4 ($CH(CH_3)_2$), 32.0 ($C(CH_3)_3$), 35.1 ($C(CH_3)_3$), 36.8 ($XA-C(CH_3)_2$), 106.2, 108.8, 124.7, 125.9, 128.4, 128.6, 131.8, 140.82, 144.2, 145.9, 147.5, 150.0 ($Ar-C$); anal. calc. for $C_{99}H_{128}Al_2K_2N_4O_2$: C 76.38%, H 8.73%, N 3.79%, found: C 76.46%, H 8.58%, N 3.65%.

$[K\{Ph(H)Al(NON)\}]_2$. A solution of $[K\{Al(NON)\}]_2$ (0.200 g, 0.136 mmol) in benzene (10 ml) was heated at 60 °C for 4 days without stirring. During the 4-day reaction, the solution slowly changed colour from yellow to almost colourless, and colourless crystals grew on the wall of the flask. The reaction was cooled to room temperature, the solution decanted from the flask and the colourless crystals dried in vacuo to give $[K\{Ph(H)Al(NON)\}]_2$ as a colourless powder (0.186 g, 84%). 1H NMR (400 MHz, THF- d_8 , 298 K): δ = 0.19 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 0.72 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 0.95 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.10 (s, 36 H, $C(CH_3)_3$), 1.25 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.69 (s, 6 H, $C(CH_3)_2$), 1.83 (s, 6 H, $C(CH_3)_2$), 3.02 (sept., $^3J_{HH}$ = 6.8 Hz, 4 H, $CH(CH_3)_2$), 3.72 (sept., $^3J_{HH}$ = 6.8 Hz, 4 H, $CH(CH_3)_2$), 5.66 (d, $^4J_{HH}$ = 1.9 Hz, 4 H, $XA-o-CH$), 6.43 (d, $^4J_{HH}$ = 1.9 Hz, 4 H, $XA-p-CH$), 6.88–7.15 (m, 18 H ArH), 7.51 (br, 4 H, ArH); $^{13}C\{^1H\}$ NMR (126 MHz, C_6D_6): δ = 23.3 ($XA-C(CH_3)_2$), 23.8, 24.9, 25.7, 26.4 ($CH(CH_3)_2$), 28.0, 28.8 ($CH(CH_3)_2$), 32.0 ($C(CH_3)_3$), 33.9 ($XA-C(CH_3)_2$), 35.2 ($C(CH_3)_3$), 36.7 ($XA-C(CH_3)_2$), 103.5, 108.8, 123.2, 124.3, 124.6, 125.2, 126.8, 128.8, 131.0, 138.1, 140.3, 145.4, 146.3, 146.4, 148.6, 150.2 ($Ar-C$); IR ν (cm^{-1} , Nujol): 1,636(s), 1,581(s), 1,415(m), 1,360(m), 1,330(m), 1,310(s), 1,254(s), 1,210(s), 1,112(m), 1,016(m), 903(m), 806(m), 781(s), 725(s), 668(s).

(NON)GaI. To a solution of $K_2(NON)$ (2.00 g, 2.67 mmol) in toluene (30 ml) was added a solution of Gal_3 (1.20 g, 2.67 mmol) in toluene (30 ml) at $-78^\circ C$ for 15 min. The reaction mixture was slowly warmed to room temperature where it was stirred for a further 16 h. The reaction mixture was filtered and volatiles from the filtrate were removed in vacuo to give an off-white solid. This solid was washed with hexane (2×20 ml) to give (NON)GaI as a colourless powder (1.80 g, 78% yield). X-ray quality crystals of (NON)GaI were obtained by recrystallizing this solid from warm toluene. 1H NMR (400 MHz, C_6D_6 , 298 K): δ = 1.10 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.16 (s, 18 H, $C(CH_3)_3$), 1.34 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.58 (s, 6 H, $XA-C(CH_3)_2$), 3.48 (sept., $^3J_{HH}$ = 6.8 Hz, 4 H, $CH(CH_3)_2$), 6.45 (d, $^4J_{HH}$ = 1.7 Hz, 2 H, $XA-o-CH$), 6.83 (s, $^4J_{HH}$ = 1.7 Hz, 2 H, $XA-p-CH$), 7.22–7.29 (m, 6 H, ArH); $^{13}C\{^1H\}$ NMR (126 MHz, C_6D_6): δ = 24.5, 25.5 ($CH(CH_3)_2$), 27.3 ($XA-CH_3$), 29.3 ($CH(CH_3)_2$), 31.7 ($C(CH_3)_3$), 35.2 ($C(CH_3)_3$), 38.0 ($XA-C(CH_3)_2$), 109.2, 112.4, 124.7, 128.0, 128.2, 128.4, 134.3, 138.7, 140.7, 141.2, 147.9, 148.9 ($Ar-C$); anal. calc. for $C_{47}H_{62}GaIN_2O$: C 65.06%, H 7.20%, N 3.23%, found: C 65.23%, H 7.36%, N 3.36%.

$[K\{Ga(NON)\}]_2$. To a suspension of KC_8 (0.103 g, 0.762 mmol) in toluene (10 ml) was added a solution of (NON)GaI (0.300 g, 0.346 mmol) in toluene (10 ml) at room temperature. The reaction mixture was stirred for 16 h at room temperature before filtration. The filtrate was concentrated in vacuo to around 5 ml and slowly cooled to $-30^\circ C$ overnight to give $[K\{Ga(NON)\}]_2$ as yellow/orange crystals (0.178 g, 66%). 1H NMR (400 MHz, C_6D_6 , 298 K): δ = 1.04 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.11 (d, $^3J_{HH}$ = 6.8 Hz, 12 H, $CH(CH_3)_2$), 1.32 (s, 18 H, $C(CH_3)_3$), 1.75 (s, 6 H, $XA-C(CH_3)_2$), 3.67 (sept., $^3J_{HH}$ = 6.8 Hz, 4 H, $CH(CH_3)_2$), 6.11 (d, $^4J_{HH}$ = 1.6 Hz, 2 H, $XA-o-CH$), 6.78 (s, $^4J_{HH}$ = 1.6 Hz, 2 H, $XA-p-CH$), 6.90–7.16 (m, 6 H, ArH); $^{13}C\{^1H\}$ NMR (101 MHz, C_6D_6): δ = 24.6, 25.7 ($CH(CH_3)_2$), 28.4 ($CH(CH_3)_2$), 28.5 ($XA-C(CH_3)_2$), 32.2 ($C(CH_3)_3$), 35.0 ($C(CH_3)_3$), 37.2 ($XA-C(CH_3)_2$), 105.5, 108.4, 124.1, 124.3, 127.9, 128.2, 133.6, 143.1, 143.2, 146.4, 148.9, 149.5 ($Ar-C$); anal. calc. for $C_{94}H_{124}Ga_2N_4O_2$: C 72.39%, H 8.01%, N 3.59%, found: C 72.07%, H 7.85%, N 3.36%.

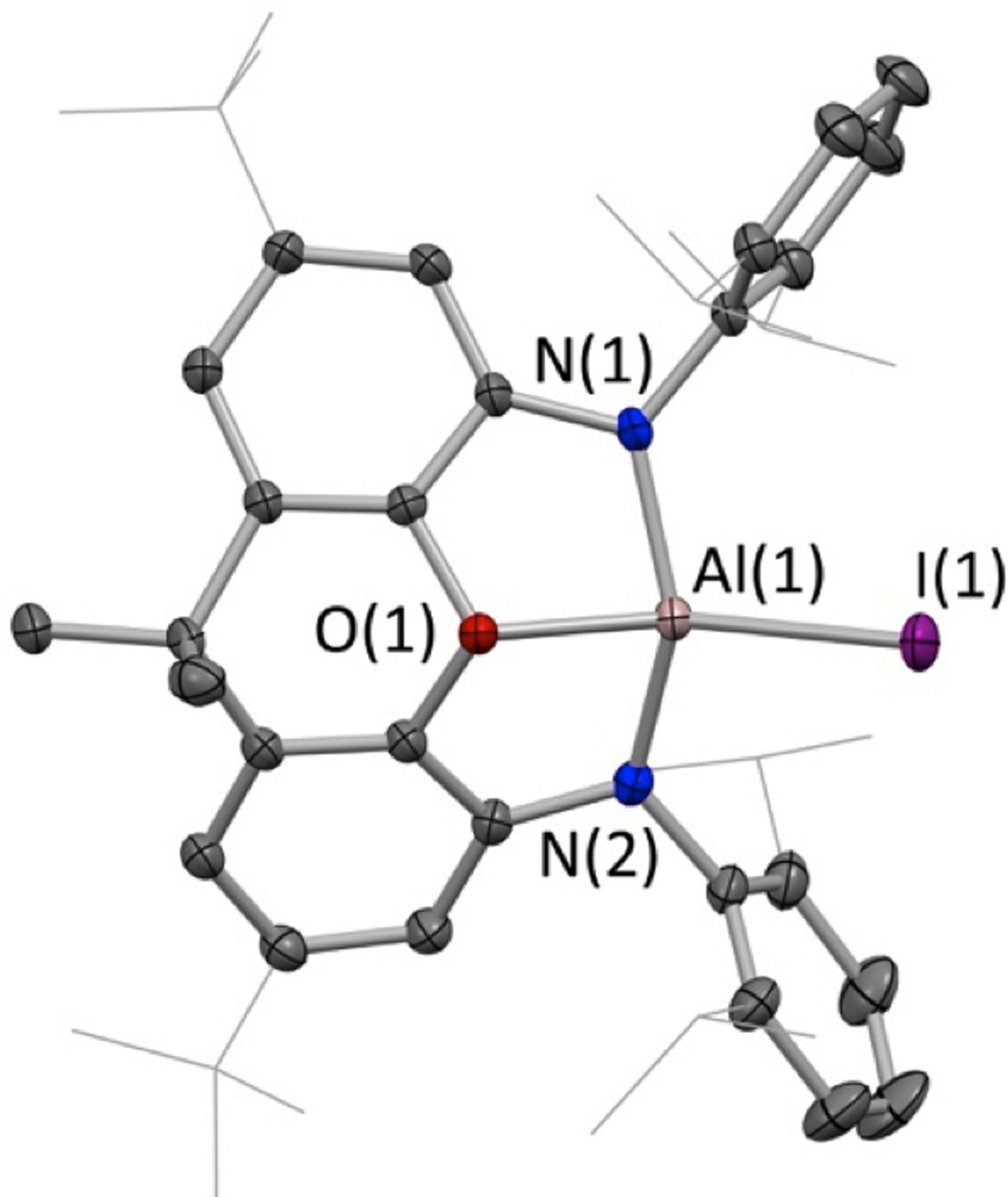
X-ray crystallographic studies. Single-crystal X-ray diffraction data were collected using an Oxford Diffraction Supernova dual-source diffractometer equipped with a 135 mm Atlas CCD area detector. Crystals were selected under Paratone-N oil, mounted on Micromount loops and quench-cooled using an Oxford Cryosystems open flow N_2 cooling device³⁰. Data were collected at 150 K (unless otherwise

stated) using either mirror monochromated $Cu K\alpha$ radiation ($\lambda = 1.5418 \text{ \AA}$; Oxford Diffraction Supernova) or $Mo K\alpha$ radiation ($\lambda = 0.71073 \text{ \AA}$; Oxford Diffraction Supernova). Data collected on the Oxford Diffraction Supernova diffractometer were processed using the CrysAlisPro package, including unit-cell parameter refinement and inter-frame scaling (which was carried out using SCALE3 ABSPACK within CrysAlisPro)³¹. Equivalent reflections were merged and diffraction patterns processed with the CrysAlisPro suite. Structures were subsequently solved using SHELXT 2014 and refined on F^2 using the SHELXL 2014 package and ShelXle or XSeed^{32,33}.

Computational details. All computations reported here were performed at the DFT level with Gaussian09 (Revision D.01) program package³⁴. Geometry optimizations were performed with the PBE1PBE exchange correlation functional^{35–37}, using def-TZVP basis sets³⁸, with Grimme's empirical dispersion correction (DFT-D3)^{39,40}. The natural bond orbital (NBO) analyses were performed using NBO 5.9 as implemented in Gaussian09⁴¹. Graphics were created with the program GaussView⁴². The geometry optimization calculations were performed for model systems ('Bu and 'Pr groups replaced by Me) in the case of the dimer $[K\{Al(NON')\}]_2$ to reduce computational cost, and full ligand systems were used in the calculations of the monomers $[Al(NON')]^-$ and $Al(Nacnac)^{Dipp}$. In addition, $[Al(NON')]^-$ and $Al(Nacnac)^{Dipp}$ were optimized in the C_1 point group and the dimer $[K\{Al(NON')\}]_2$ in the C_i point group. The nature of stationary points found (minimum) was in most cases confirmed by full frequency calculations; the optimized structure of the dimer $[K\{Al(NON')\}]_2$, however, has one imaginary frequency of only $9i \text{ cm}^{-1}$ corresponding to the vibration of the whole system. Unfortunately, we were unable to find a global minimum with several optimization attempts (using tighter convergence criteria, no symmetry, or different starting geometries). However, if the dimer was optimized without the empirical dispersion correction, a minimum could be found, but the energy of the system was found to be almost 600 kJ mol^{-1} higher in energy compared to the $[K\{Al(NON')\}]_2$ optimized using dispersion correction. This highlights the importance of dispersion in these systems. Consequently, we performed the NBO analysis to the structure obtained from optimization with dispersion correction.

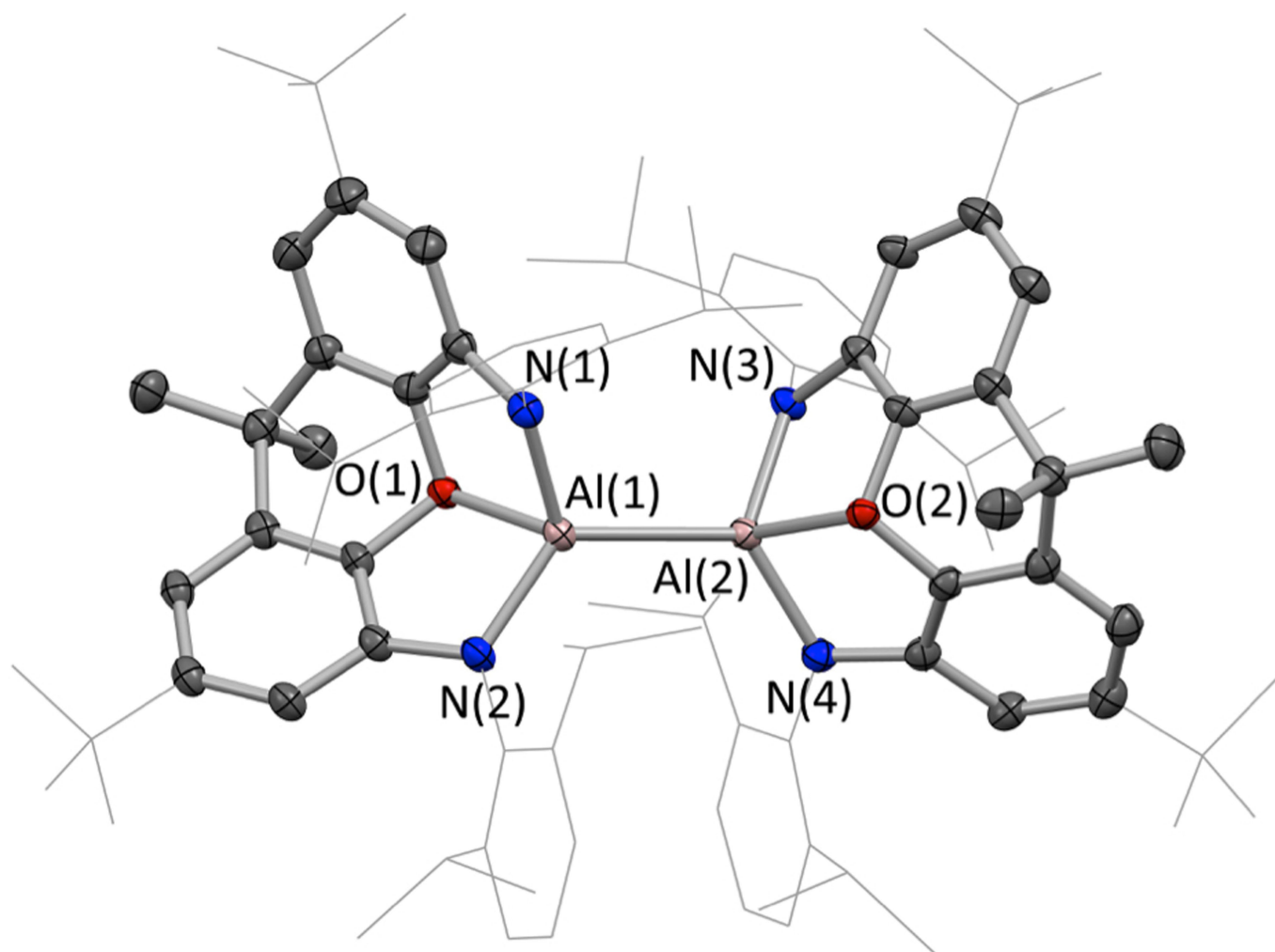
Data availability. The spectroscopic data that support the findings of this study are available from the corresponding authors upon reasonable request. X-ray crystallographic data for (NON)AlI, $[Al(NON)]_2$, $[K\{Al(NON)\}]_2$, $[K\{H_2Al(NON)\}]_2$, $[K\{Ph(H)Al(NON)\}]_2$, (NON)AlH, (NON)AlMe, (Nacnac)^{Me}MgAl(NON), (NON)GaI and $[K\{Ga(NON)\}]_2$ are available in the Supplementary Information and from the Cambridge Crystallographic Data Centre (<https://www.ccdc.cam.ac.uk/>) under reference numbers 1581591–1581600.

- Cruz, C. A., Emslie, D. J. H., Harrington, L. E., Britten, J. F. & Robertson, C. M. Extremely stable thorium(IV) dialkyl complexes supported by rigid tridentate 4,5-bis(anilido)xanthene and 2,6-bis(anilidomethyl)pyridine ligands. *Organometallics* **26**, 692–701 (2007).
- Bonyhady, S. J. et al. β -diketiminate-stabilized magnesium(I) dimers and magnesium(II) hydride complexes: synthesis, characterization, adduct formation, and reactivity studies. *Chem. Eur. J.* **16**, 938–955 (2010).
- Zakharkin, L. I. & Gavrilenco, V. V. Mutual conversions in the aluminohydrides of lithium, sodium, and potassium. *Russ. Chem. Bull.* **11**, 1076–1078 (1962).
- Cosier, J. & Glazer, A. M. A nitrogen-gas-stream cryostat for general X-ray diffraction studies. *J. Appl. Crystallogr.* **19**, 105–107 (1986).
- CrysAlisPro v.1.171.35.8 (Agilent Technologies, 2011).
- Sheldrick, G. M. SHELX-2014 (2014).
- Hübschle, C. B., Sheldrick, G. M. & Dittrich, B. *ShelXle*: a Qt graphical user interface for SHELXL. *J. Appl. Crystallogr.* **44**, 1281–1284 (2011).
- Frisch, M. J. et al. Gaussian 09 Rev. D.01. (Gaussian Inc., 2009).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- Perdew, J. P., Ernzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **105**, 9982–9985 (1996).
- Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999).
- Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100**, 5829–5835 (1994).
- Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104–154119 (2010).
- Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
- Glendening, E. D. et al. NBO v. 5.9 (2011).
- Dennington, R., Keith, T. A. & Millam, J. M. GaussView v. 5.0 (Semichem Inc., 2009).



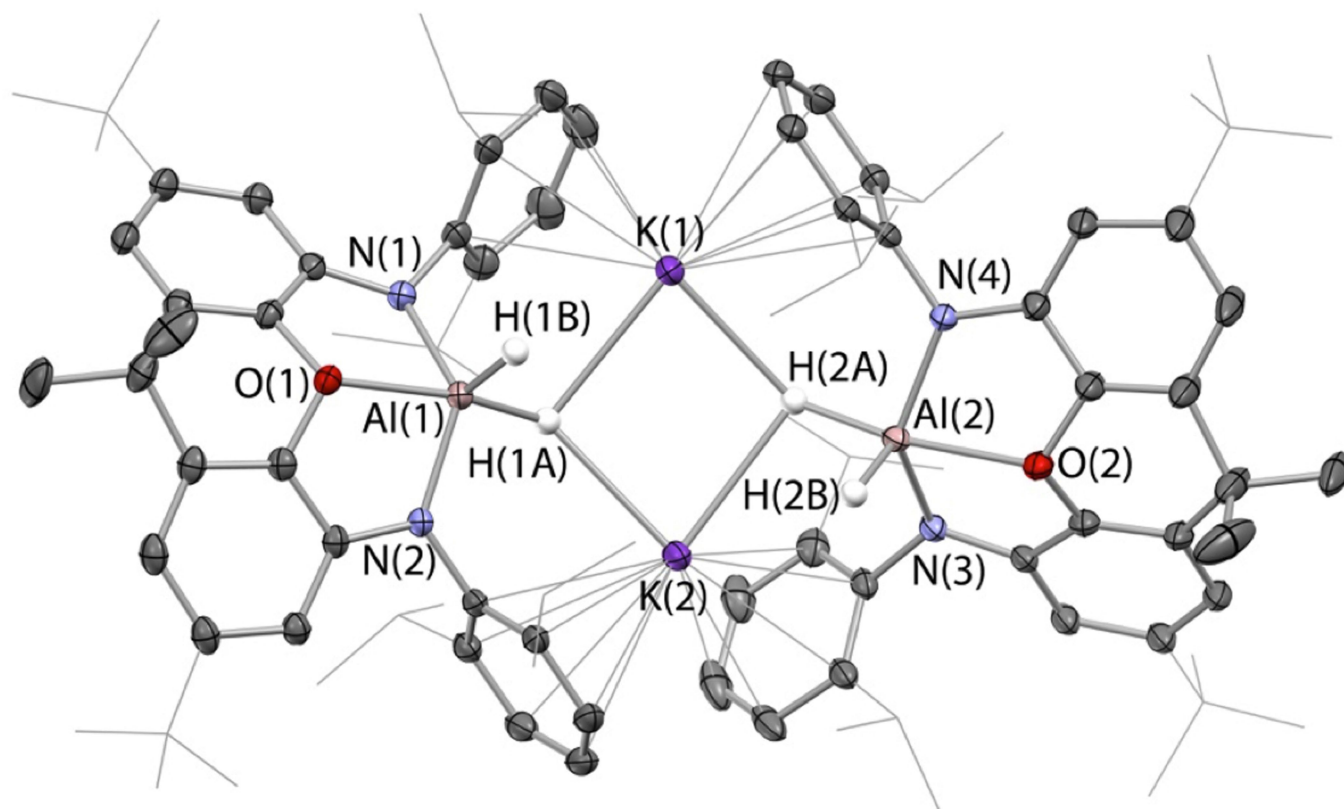
Extended Data Fig. 1 | Molecular structure of (NON)AlI as determined by X-ray crystallography. Hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity; thermal

ellipsoids have been drawn at the 35% probability level. Key bond lengths (Å) and angles (°): Al(1)–I(1) 2.497(1), Al(1)–N(1) 1.846(2), Al(1)–N(2) 1.846(2), Al(1)–O(1) 1.967(2), N(1)–Al(1)–N(2) 143.0(1).



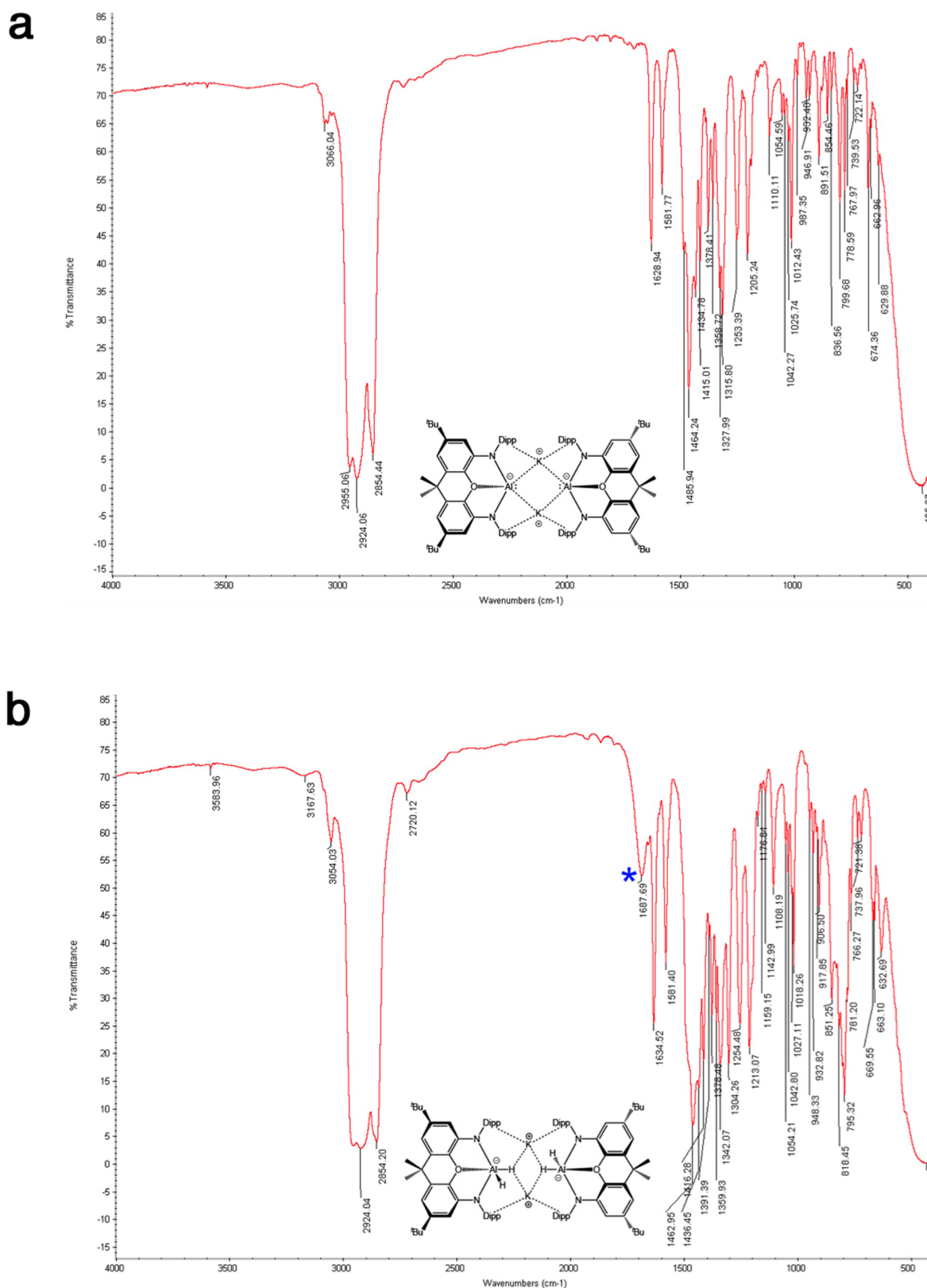
Extended Data Fig. 2 | Molecular structure of $[\text{Al}(\text{NON})_2]_2$ as determined by X-ray crystallography. Hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity; thermal ellipsoids have been drawn at the 35% probability level.

Key bond lengths (Å) and angles (°): Al(1)–Al(2) 2.646(1), Al(1)–N(1) 1.902(2), Al(1)–N(2) 1.895(2), Al(2)–N(3) 1.901(2), Al(2)–N(4) 1.900(2), Al(1)–O(1) 1.976(2), Al(2)–O(2) 1.981(2), N(1)–Al(1)–N(2) 119.0(1), N(3)–Al(2)–N(4) 118.6(1).



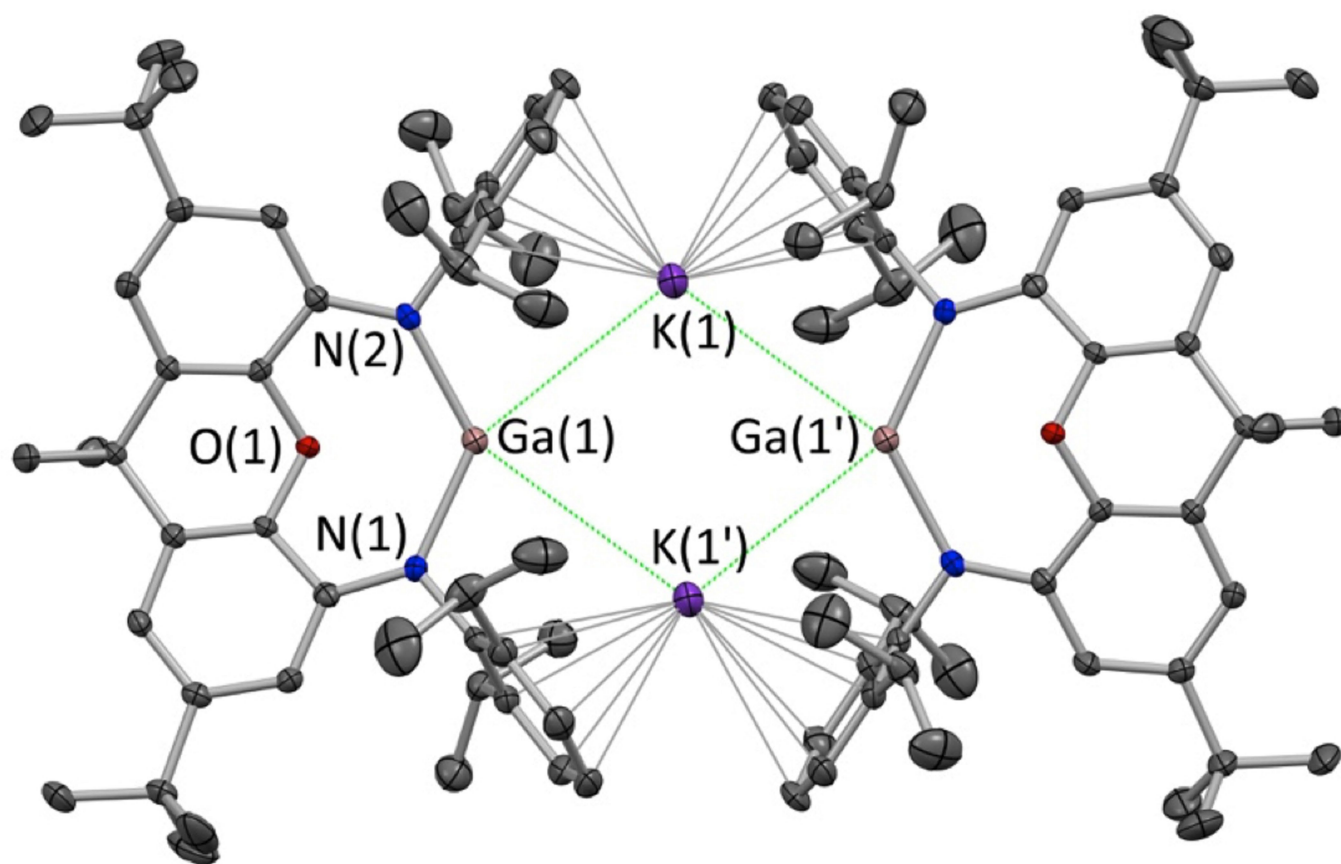
Extended Data Fig. 3 | Molecular structure of one of the molecules in the asymmetric unit of $[K\{H_2Al(NON)\}]_2$ as determined by X-ray crystallography. Second (essentially identical) component, benzene solvate molecules and carbon-bound hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity; thermal ellipsoids have been drawn at the 35% probability level. Key distances

(Å) and angles ($^\circ$): Al(1)–N(1/2) 1.933(2)/1.921(2), Al(2)–N(3/4) 1.934(2)/1.917(2), Al(1)–O(1) 2.131(1), Al(2)–O(2) 2.124(2), Al(1)–N(2) 6.356(1), Al(1)–K(1/2) 3.648(1)/4.065(1), Al(2)–K(1/2) 3.580(1)/4.039(1), Al(1)–H(1A/1B) 1.69(4)/1.55(4), Al(2)–H(2A/2B) 1.71(4)/1.58(4), N(1)–Al(1)–N(2) 130.3(1), N(3)–Al(2)–N(4) 131.1(1).



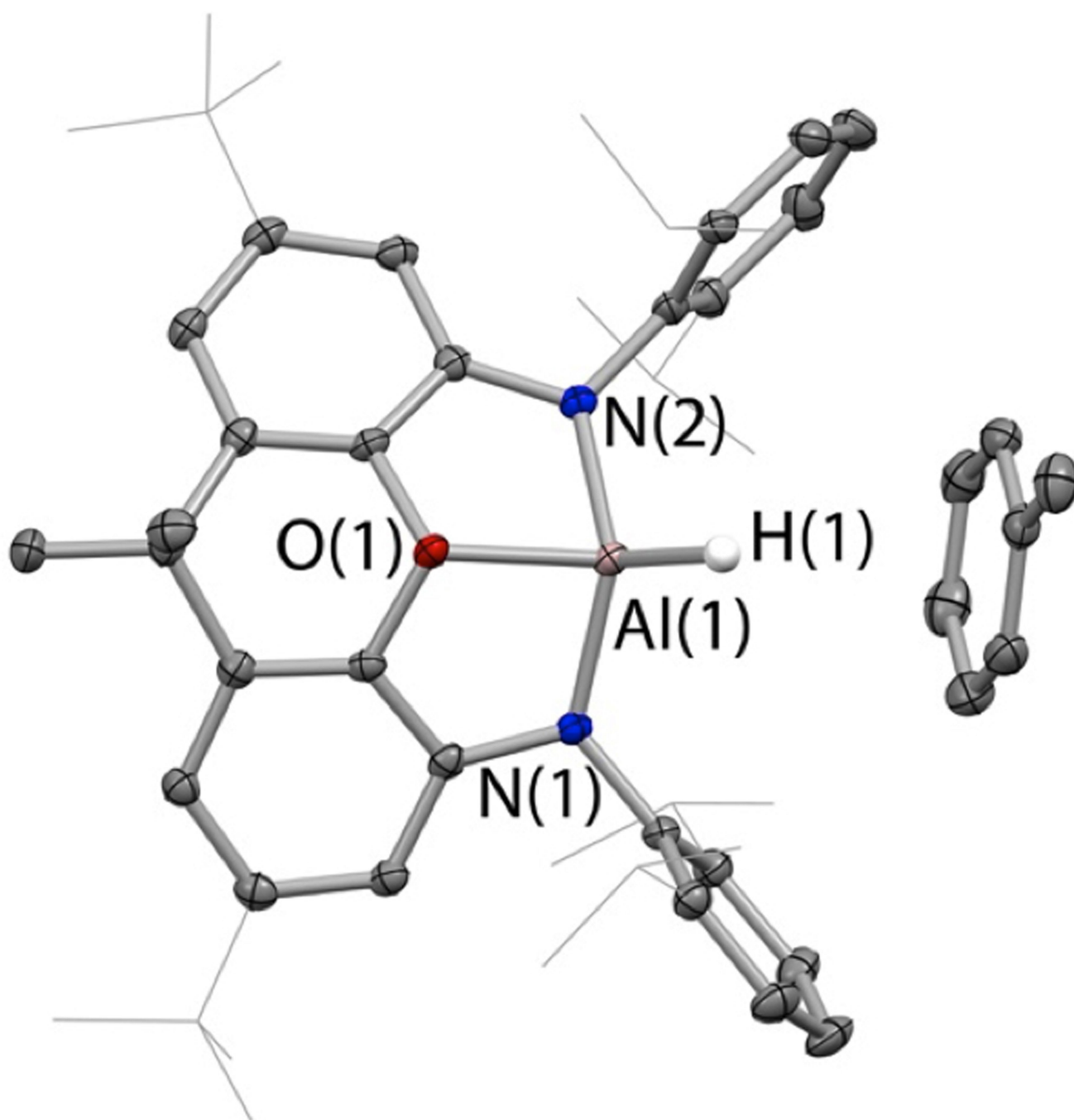
Extended Data Fig. 4 | Infrared spectra of $[K\{Al(NON)\}]_2$ and $[K\{H_2Al(NON)\}]_2$. a, $[K\{Al(NON)\}]_2$. b, $[K\{H_2Al(NON)\}]_2$. Both spectra

have been measured on samples as Nujol mulls; the blue asterisk highlights the Al–H stretching band of $[K\{H_2Al(NON)\}]_2$.



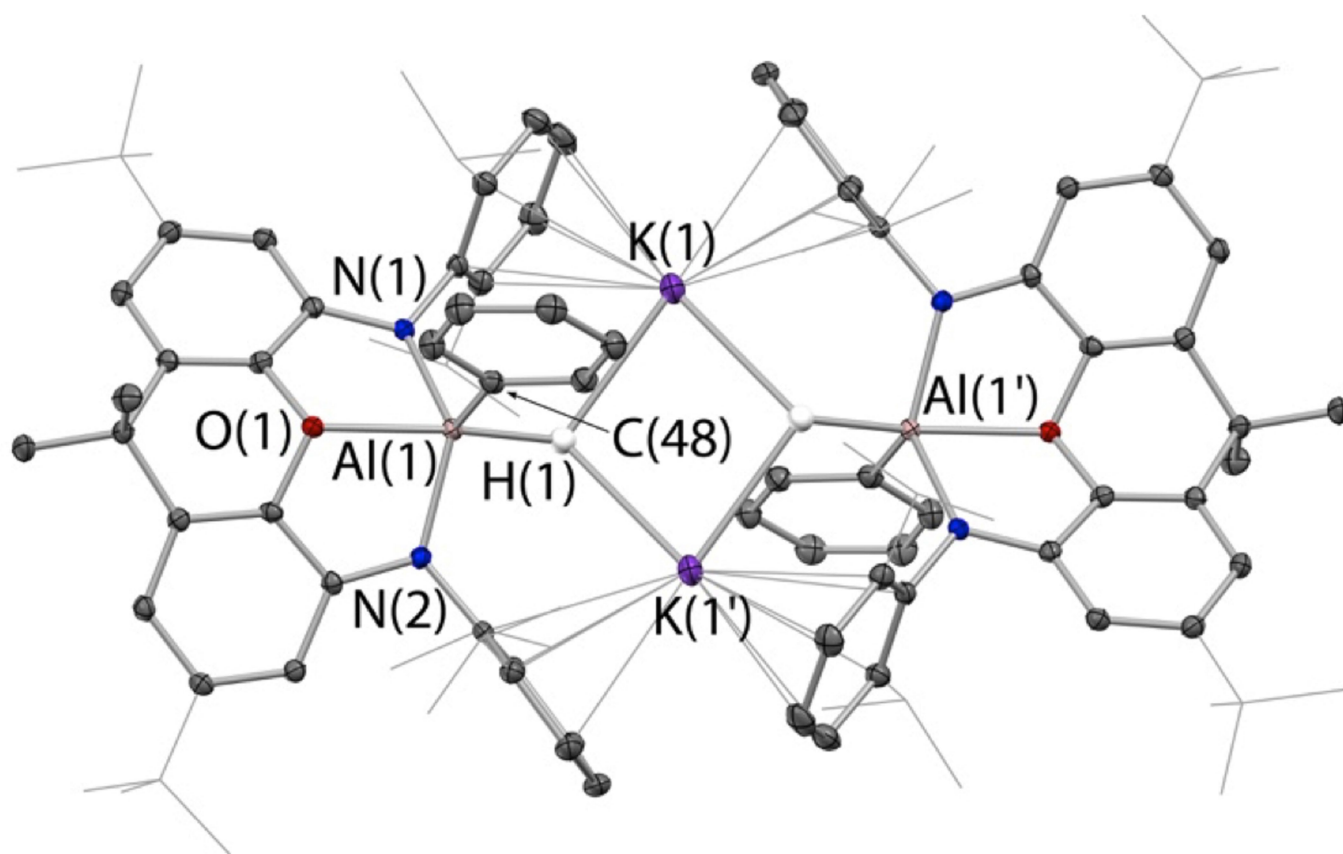
Extended Data Fig. 5 | Molecular structure of $[K\{Ga(NON)\}]_2$ as determined by X-ray crystallography. Hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity; thermal ellipsoids have been drawn at the 35% probability level. Key

bond lengths and distances (Å) and angles (°): Ga(1) ... Ga(1') 6.134(1), Ga(1) ... K(1) 3.970(1), Ga(1) ... K(1') 3.784(1), Ga(1)–N(1) 2.093(2), Ga(1)–N(2) 2.106(2), Ga(1)–O(1) 2.542(2), N(1)–Ga(1)–N(2) 126.0(1).



Extended Data Fig. 6 | Molecular structure of (NON)AlH·toluene as determined by X-ray crystallography. Most hydrogen atoms have been omitted and selected carbon atoms shown in wireframe format for clarity;

thermal ellipsoids have been drawn at the 35% probability level. Key bond lengths (Å) and angles (°): Al(1)–N(1) 1.873(1), Al(1)–N(2) 1.872(1), Al(1)–O(1) 1.944(1), Al(1)–H(1) 1.49(2), N(1)–Al(1)–N(2) 134.1(1).



Extended Data Fig. 7 | Molecular structure of $[K\{Ph(H)Al(NON)\}]_2$ as determined by X-ray crystallography. Most hydrogen atoms and benzene solvate molecules have been omitted, and selected carbon atoms shown in wireframe format for clarity; thermal ellipsoids have been drawn at the

35% probability level. Key bond lengths (Å) and angles (°): Al(1)–N(1) 1.945(2), Al(1)–N(2) 1.944(2), Al(1)–O(1) 2.122(1), Al(1)–C(48) 2.007(1), Al(1)–H(1) 1.82(3), N(1)–Al(1)–N(2) 132.2(1).

Extended Data Table 1 | Selected X-ray data collection and refinement parameters for complexes prepared in this study

	(NON)AlI-toluene	[Al(NON) ₂]-2hexane-pentane	[K ₂ Al(NON)] ₂ -2benzene	(NON)AlI-toluene
formula	C ₁₀ H ₁₅ AlIN ₂ O	C ₁₁ H ₁₄ AlN ₂ O ₂	C ₁₃ H ₁₄ Al ₂ K ₂ N ₂ O ₂	C ₁₀ H ₁₅ AlN ₂ O
fw [g mol ⁻¹]	917.00	1640.41	1786.56	791.10
crystal system	monoclinic	monoclinic	monoclinic	monoclinic
space group	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> 2 ₁ / <i>c</i>	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> 2 ₁ / <i>n</i>
<i>a</i> (Å)	15.5890(4)	17.5615(2)	12.7464(4)	15.3319(2)
<i>b</i> (Å)	16.6528(5)	18.5034(2)	24.0329(8)	18.4919(2)
<i>c</i> (Å)	19.5742(4)	33.8933(5)	18.0006(5)	17.1480(2)
α (°)	90	90	90	90
β (°)	101.021(2)	102.156(1)	101.359(3)	104.980(2)
γ (°)	90	90	90	90
<i>V</i> (Å ³)	4987.8(2)	10766.6(2)	5406.2(3)	4696.51(11)
<i>Z</i>	4	4	2	4
radiation, λ (Å)	Cu K α (1.54184)	Cu K α (1.54184)	Cu K α (1.54184)	Cu K α (1.54184)
<i>T</i> (K)	175(2)	200(2)	100(2)	150(2)
ρ_{calc} (g cm ⁻³)	1.221	1.012	1.098	1.119
μ (mm ⁻¹)	5.491	0.588	1.304	0.662
reflections collected	57769	64900	31957	27914
independent reflections	10426	22254	11147	9730
parameters	580	1256	696	544
<i>R</i> _{int}	0.0501	0.0325	0.0308	0.0273
<i>R</i> 1 (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	3.84/5.10	7.40/8.91	7.22/8.14	3.98/4.75
<i>wR</i> 2(<i>F</i> ²) (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	9.62/10.61	20.83/22.32	18.23/19.10	10.18/10.80
GOF	1.017	1.086	1.038	1.030
CCDC deposition number	1581591	1581595	1581592	1581598

	(NON)AlMe-1.5toluene	(Nacnac) ^{Ni} MgAl(NON)	[K ₂ H ₂ Al(NON)] ₂ -3.25toluene
formula	C _{10.5} H _{17.5} AlN ₂ O	C ₁₀ H ₁₁ AlMgN ₂ O	C _{13.5} H _{17.5} Al ₂ K ₂ N ₄
fw [g mol ⁻¹]	851.20	1055.75	1732.01
crystal system	triclinic	monoclinic	triclinic
space group	<i>P</i> -1	<i>P</i> 2 ₁ / <i>c</i>	<i>P</i> -1
<i>a</i> (Å)	11.3959(3)	22.3166(5)	19.2554(3)
<i>b</i> (Å)	19.1112(4)	13.6123(3)	20.2365(5)
<i>c</i> (Å)	26.1596(4)	22.5503(5)	28.1461(4)
α (°)	99.313(2)	90	86.636(2)
β (°)	96.019(2)	108.156(2)	89.945(1)
γ (°)	106.398(2)	90	72.808(2)
<i>V</i> (Å ³)	5324.9(2)	6509.3(3)	10457.8(4)
<i>Z</i>	4	4	4
radiation, λ (Å)	Cu K α (1.54184)	Cu K α (1.54184)	Cu K α (1.54184)
<i>T</i> (K)	200(2)	150(2)	100(2)
ρ_{calc} (g cm ⁻³)	1.062	1.077	1.100
μ (mm ⁻¹)	0.614	0.687	1.333
reflections collected	61949	39799	118805
independent reflections	22050	13468	43335
parameters	1439	748	2867
<i>R</i> _{int}	0.0433	0.0505	0.0295
<i>R</i> 1 (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	6.89/8.57	5.35/6.63	5.77/7.06
<i>wR</i> 2(<i>F</i> ²) (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	19.12/21.49	13.76/15.20	15.43/16.73
GOF	1.032	1.035	1.017
CCDC deposition number	1581597	1581593	1581600

	[K ₂ (Ph(H)Al(NON))] ₂ -6benzene	(NON)GaI-toluene	[K ₂ (Ga(NON))] ₂ -4toluene
formula	C ₁₀₂ H ₁₂₂ Al ₂ K ₂ N ₂ O ₂	C ₁₀ H ₁₅ GaN ₂ O	C ₁₂₂ H ₁₁₆ Ga ₂ K ₂ N ₂ O ₂
fw [g mol ⁻¹]	2098.99	959.74	1928.14
crystal system	triclinic	monoclinic	triclinic
space group	<i>P</i> -1	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> -1
<i>a</i> (Å)	13.9496(7)	15.5521(6)	12.3217(5)
<i>b</i> (Å)	14.3853(7)	16.6088(9)	14.6039(7)
<i>c</i> (Å)	17.5433(8)	19.5714(7)	17.8128(7)
α (°)	89.002(4)	90	71.255(4)
β (°)	71.939(4)	100.820(3)	87.125(4)
γ (°)	67.043(5)	90	65.757(4)
<i>V</i> (Å ³)	3060.2(3)	4965.5(4)	2755.2(2)
<i>Z</i>	1	4	1
radiation, λ (Å)	Cu K α (1.54184)	Cu K α (1.54184)	Mo K α (0.71073)
<i>T</i> (K)	100(2)	150(2)	150(2)
ρ_{calc} (g cm ⁻³)	1.139	1.284	1.162
μ (mm ⁻¹)	1.223	5.916	
reflections collected	32283	29243	31412
independent reflections	12661	10229	14357
parameters	704	549	656
<i>R</i> _{int}	0.0259	0.0576	0.0456
<i>R</i> 1 (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	4.64/5.63	4.98/6.25	5.45/9.60
<i>wR</i> 2(<i>F</i> ²) (<i>I</i> ≥ 2 σ (<i>I</i>)/all data)	12.16/12.93	12.61/13.87	9.97/11.82
GOF	1.041	1.036	1.022
CCDC deposition number	1581599	1581596	1581594

Complete *Ichthyornis* skull illuminates mosaic assembly of the avian head

Daniel J. Field^{1,6,7}, Michael Hanson^{1,7}, David Burnham², Laura E. Wilson³, Kristopher Super², Dana Ehret⁴, Jun A. Ebersole⁵ & Bhart-Anjan S. Bhullar^{1*}

The skull of living birds is greatly modified from the condition found in their dinosaurian antecedents. Bird skulls have an enlarged, toothless premaxillary beak and an intricate kinetic system that includes a mobile palate and jaw suspensorium. The expanded avian neurocranium protects an enlarged brain and is flanked by reduced jaw adductor muscles. However, the order of appearance of these features and the nature of their earliest manifestations remain unknown. The Late Cretaceous toothed bird *Ichthyornis dispar* sits in a pivotal phylogenetic position outside living groups: it is close to the extant avian radiation but retains numerous ancestral characters^{1–3}. Although its evolutionary importance continues to be affirmed^{3–8}, no substantial new cranial material of *I. dispar* has been described beyond incomplete remains recovered in the 1870s. Jurassic and Cretaceous Lagerstätten have yielded important avialan fossils, but their skulls are typically crushed and distorted⁹. Here we report four three-dimensionally preserved specimens of *I. dispar*—including an unusually complete skull—as well as two previously overlooked elements from the Yale Peabody Museum holotype, YPM 1450. We used these specimens to generate a nearly complete three-dimensional reconstruction of the *I. dispar* skull using high-resolution computed tomography. Our study reveals that *I. dispar* had a transitional beak—small, lacking a palatal shelf and restricted to the tips of the jaws—coupled with a kinetic system similar to that of living birds. The feeding apparatus of extant birds therefore evolved earlier than previously thought and its components were functionally and developmentally coordinated. The brain was relatively modern, but the temporal region was unexpectedly dinosaurian: it retained a large adductor chamber bounded dorsally by substantial bony remnants of the ancestral reptilian upper temporal fenestra. This combination of features documents that important attributes of the avian brain and palate evolved before the reduction of jaw musculature and the full transformation of the beak.

The adaptability of the avian skull is manifest in its considerable functional disparity across living birds¹⁰. However, the earliest appearances of several avian cranial innovations and their sequence of acquisition are incompletely known, owing to the uneven and largely two-dimensional preservation of early ornithuran cranial material^{1,3,11,12}. We report a newly discovered and nearly complete skull of the iconic early ornithuran *I. dispar* housed at the Sternberg Museum of Natural History (FHSM). One of us (K.S.) discovered the specimen (FHSM VP-18702) in 2014 near Castle Rock in Gove County (Kansas); it derives from the base of lithostratigraphic marker unit 10 in the Smoky Hill Member of the Niobrara Formation (Middle Santonian stage, Late Cretaceous).

We have assembled an almost complete three-dimensional reconstruction of the skull of *I. dispar*¹³ (Figs. 1, 2, Extended Data Figs. 1–4 and Supplementary Information 1), using the FHSM VP-18702 specimen, the holotype (YPM 1450) as well as elements from three other

undescribed specimens (from the Alabama Museum of Natural History (ALMNH 3316), the University of Kansas Biodiversity Institute (KUPV 119673) and the Black Hills Institute of Geological Research (BHI 6421)), all of which we refer to *I. dispar* using multiple postcranial autapomorphies (Supplementary Fig. 1). The skull of *I. dispar* (Figs. 1, 2) illustrates a transitional point in the evolutionary history of birds stemward of Hesperornithiformes, a phylogenetic position that, after adding new characters derived from the present study, we recover consistently with the majority of recent analyses^{11,14} (Fig. 3a; see Supplementary Information 1, 2).

The upper margin of the beak is concave in profile, which is a derived condition that is shared by living birds and *I. dispar* but not by stemward ornithuromorphs¹¹ (Fig. 3b, Extended Data Figs. 5–10 and Supplementary Videos 1, 5). As in Aves, the fused, toothless premaxillae are acutely pointed and have a terminal hook. They occupy only the anteriormost quarter of the rostrum; short premaxillae restricted to the tip of the snout also characterize stemward ornithuromorphs, such as *Gansus* and *Iteravis*, and therefore appear to have been the primitive form of the avian beak^{15,16} (Fig. 3b). Neurovascular foramina indicate the presence of a highly keratinized region of rhamphotheca called the premaxillary nail, which would have enveloped and accentuated the terminal hook¹⁷. Osteological correlates for the additional rhamphothecal plates of extant birds are absent¹⁷; therefore, we infer that the nail alone was the original keratinized beak. The ventral surface of the premaxillae¹⁸ is vaulted dorsally with a median ridge for a soft-tissue internarial septum (Figs. 1, 2, 3b, Extended Data Figs. 6, 8 and Supplementary Videos 1, 5), which represents a plesiomorphic dinosaurian condition that is dissimilar to that of Hesperornithiformes¹⁹ and Aves²⁰, in which a flat palatal shelf is covered by rhamphotheca²¹. The posterior half of the palatal surface is dimpled by three pits on each side for lower teeth (Fig. 4b). The portion of the premaxillae rostral to these pits would have interacted in a pincer-like fashion with a similarly pointed predentary bone¹³.

Previously known only from a small fragment, the maxilla is plesiomorphically long (Figs. 1–3 and Supplementary Videos 1, 7). Teeth occupy sockets along the entire length of the maxilla with no indication of reduction in size relative to stemward Avialae; rows of pits run along the palate, medial to these teeth, to accommodate the mandibular dentition. Interdental ossifications develop through ontogeny²², which represents a primitive condition that is lost in Hesperornithiformes, in which small teeth are set in a continuous sulcus^{19,23} (Fig. 3b). The maxillae are robust in lateral view compared to those of crown birds, in which these bones are reduced to flat, predominantly palatal elements^{24,25}. However, *I. dispar* shares with more crownward avialans extensive maxillary shelves that form a bony palate¹⁹ (Figs. 1, 2 and Supplementary Video 1).

We discovered a substantially complete lacrimal and nasal in material composing the holotype, YPM 1450³ (Figs. 1, 2, Extended Data Fig. 10 and Supplementary Video 4). The lacrimal, which is rarely intact in

¹Department of Geology & Geophysics and Peabody Museum of Natural History, Yale University, New Haven, CT, USA. ²Biodiversity Institute and Natural History Museum, University of Kansas, Lawrence, KS, USA. ³Sternberg Museum of Natural History and Department of Geosciences, Fort Hays State University, Hays, KS, USA. ⁴Alabama Museum of Natural History, Tuscaloosa, AL, USA. ⁵McWane Science Center, Birmingham, AL, USA. ⁶Department of Biology & Biochemistry, Milner Centre for Evolution, University of Bath, Bath, UK. ⁷These authors contributed equally: Daniel J. Field, Michael Hanson. *e-mail: bhart-anjan.bhullar@yale.edu

Ichthyornis dispar (Marsh, 1872)
Skull (composite of multiple specimens)

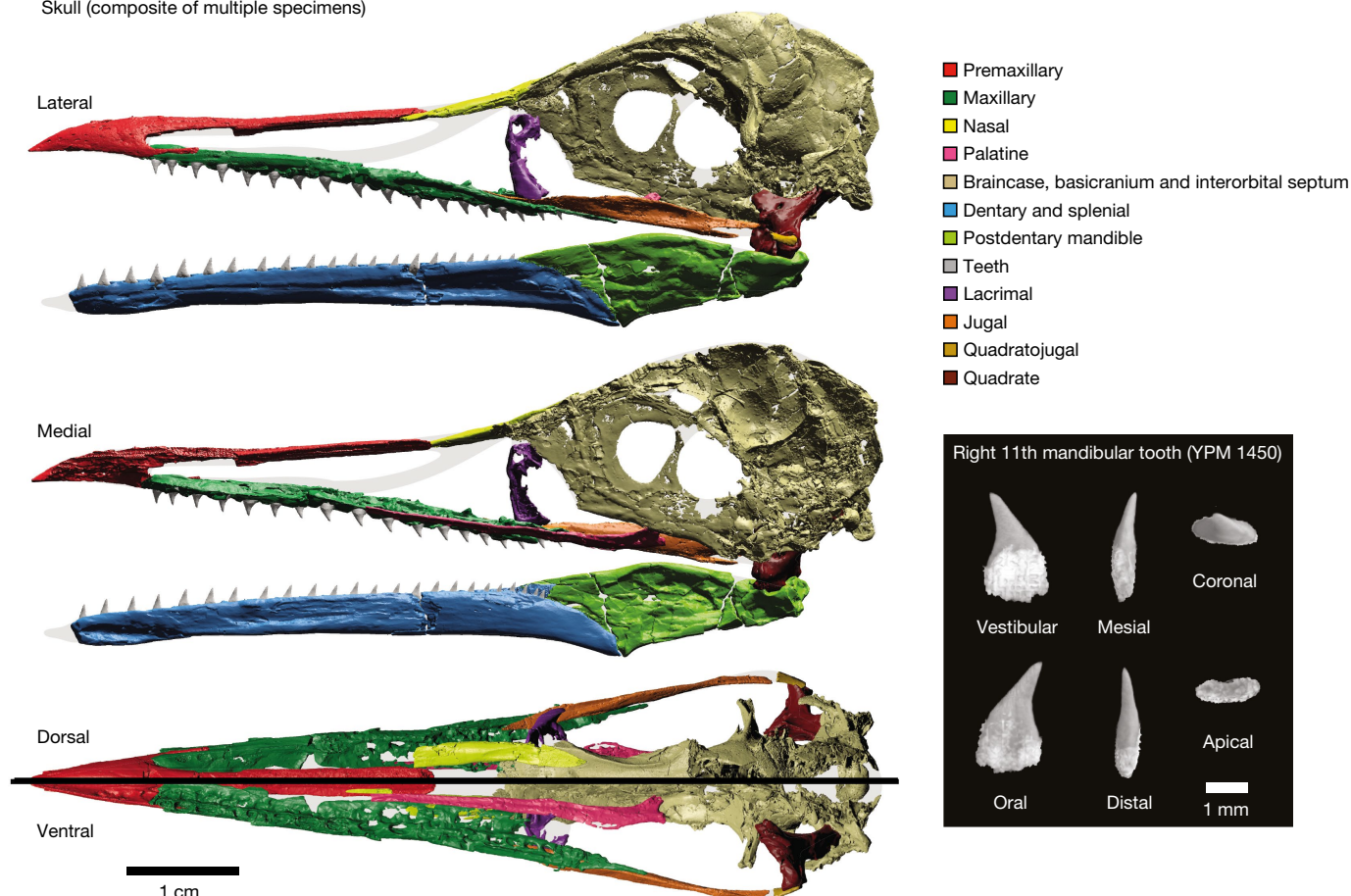


Fig. 1 | Full 3D reconstruction of the skull of *I. dispar*. A full 3D reconstruction of the skull of *I. dispar* is shown on the left. High-resolution scans of the right 11th mandibular tooth of YPM 1450 are shown on the right.

fossil avialans, is similar to that of crown-clade birds^{1,9} in being curved caudally and perforated by a capacious lacrimal foramen. The frontals, nasal and premaxillae interdigitated with one another on the dorsal surface of the skull such that—similar to palaeognath birds and unlike neognath birds—*I. dispar* lacked a transverse naso-frontal hinge^{20,26–28}. A sulcus on the rostral half of the maxilla (Figs. 1, 2 and Extended Data Fig. 1) indicates a broad naso-maxillary contact and a correspondingly broad postnarial bar²⁷. This condition resembles that in Neognathae, which suggests that the narrow or absent contact and bar of the Palaeognathae^{20,25,29} are derived within crown birds.

The palatine, rarely preserved in Mesozoic avialans²⁶, is narrow and elongate (Figs. 1, 2, Extended Data Fig. 6 and Supplementary Video 1), and is similar to that of crown birds in being unsutured to the maxilla¹⁹. The next-most-crownward examples of interpretable palatines are from the enantiornithine *Gobipteryx minuta*²⁶, in which they are broad and flat with extensive maxillary sutures^{14,30}.

The crown-bird-like palatine of *I. dispar* was linked by the unpreserved pterygoid to a quadrate that is essentially indistinguishable from that of crown birds (Figs. 1, 2, Extended Data Figs. 5–7, 10 and Supplementary Video 6). As with the quadrates of certain Neognathae (Anseriformes, Columbiformes and the Cretaceous *Vegavis*), the quadrate of *I. dispar* exhibits two rounded capitular condyles that fit into cotyles on the prootic and squamosal bones to form a mobile joint with the cranium^{3,14,20,24}; this raises the possibility that a bicondylar morphology is plesiomorphic for Aves³¹. Both constituents of the quadrato-maxillary bar, the quadratojugal and jugal, are preserved in the new material (Figs. 1, 2, Extended Data Figs. 5, 6 and Supplementary Video 3). The articular surface of the quadratojugal would have formed a mobile joint with the quadrate. The jugal is deep in lateral view, and

is dissimilar to the rod-shaped jugal of most crown birds. The tandem arrangement of the rostrum, jugal, and quadratojugal, the mobile suspensorium and the narrow, linear palatine all indicate that *I. dispar* possessed a fully functional avian cranial kinetic system, the most stemward known occurrence of this key evolutionary innovation^{20,24,25}.

The almost complete postorbital cranium of the FHSM VP-18702 skull (Figs. 1, 2, 4, Extended Data Fig. 6 and Supplementary Videos 2, 3) includes a mesethmoid interorbital septum that terminates at the anterior end of the frontals, as is the case in Hesperornithiformes and Neognathae but not in Palaeognathae. Thus, the palaeognath condition, in which the mesethmoid extends forward to form part of the internarial septum, may be autapomorphic^{18,25,28}.

The endocranial cavity in sagittal section is essentially crown-like (Fig. 4a). The forebrain was enlarged and posteroventrally rotated and the optic lobes were inflated and laterally shifted, as in living birds and the putative ornithurine *Cerebavis cenomanica*^{32–34}.

The expansive upper temporal fenestra, on the other hand, is unlike that of living birds. It is enclosed almost entirely by bone and the jaw adductor muscles within would have been substantial (Figs. 1, 2, 4b). This apparently plesiomorphic configuration is similar to that of deinonychosaurs and is entirely unexpected in an avialan that is crownward of Enantiornithes. Most extant birds have a reduced adductor chamber devoid of external skeletal boundaries, although some derived neoavians that capture prey underwater—such as cormorants and penguins—have secondarily enlarged adductor attachments on the skull roof^{19,28,35}. The anterior margin of the fenestra in *Ichthyornis* is bounded by an extensive postorbital ossification. This postorbital process resembles the separate postorbital bone of non-avialan dinosaurs in extending laterally and then posteriorly (Fig. 4b

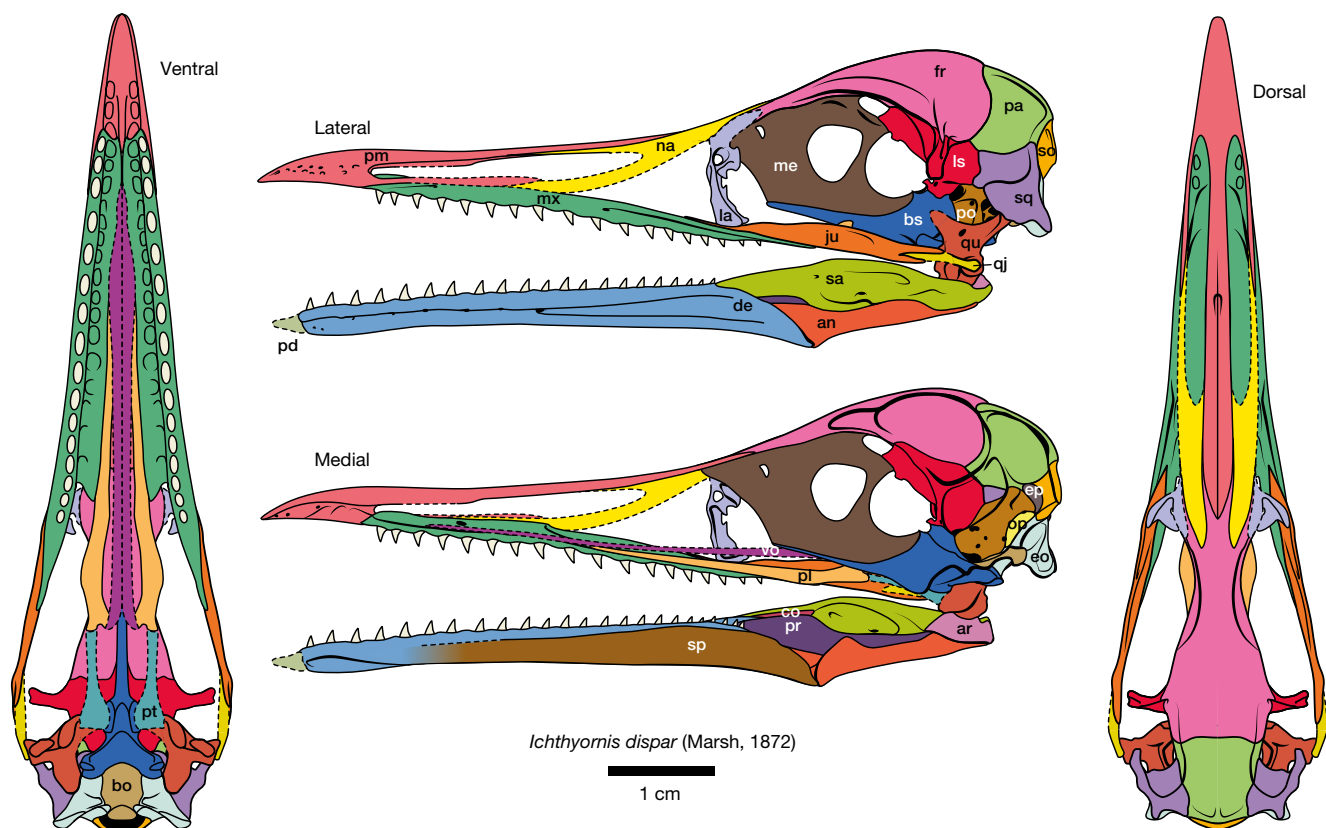


Fig. 2 | Line drawings of the skull of *I. dispar*. Solid lines indicate areas known from fossil specimens, dashed lines indicate unknown areas reconstructed from other ornithuran birds. an, angular (os angulare); ar, articular; bo, basioccipital; bs, basisphenoid; co, coronoid (os coronoideum); de, dentary (os dentale); eo, exoccipital; ep, epiotic; fr, frontal (os frontale); ju, jugal; la, lacrimal (os lacrimale); ls, laterosphenoid;

me, mesethmoid (os mesethmoidale); mx, maxilla; na, nasal (os nasale); op, opisthotic; pa, parietal; pl, palatine; pm, premaxillae; po, prootic; pr, prearticular (os prearticular); pt, pterygoid; qj, quadratojugal (os quadratojugale); qu, quadrate (os quadratum); sa, surangular (os supra-angulare); so, supraoccipital; sp, splenial (os spleniale); sq, squamosal; vo, vomer.

and Supplementary Video 2); its condition is dissimilar to that of the postorbital process of crown-clade birds, which possesses an apex that—even when secondarily enlarged or in contact with the squamosal (as in multiple highly nested clades)^{25,36}—is directed ventrally, at a right angle to the vanished upper temporal bar^{9,19,20,27,37}. Because the postorbital region of the skull is generally preserved poorly in stemward Avialae, it is unclear whether the postorbital ossification in *Ichthyornis* is secondarily enlarged—although, if so, it is enlarged in a way that is unseen in all known crown-clade birds and that is markedly convergent on non-avian theropods (Fig. 4)—or represents the late retention of an ancestral dinosaurian condition. The latter would imply evolutionary fusion of the postorbital bone to the skull roof or the retention of its shape, perhaps mediated by the persistence of its membranous embryonic precursor, despite replacement by ossified extensions from the calvarium and the laterosphenoid. The crown-bird-like contribution of the laterosphenoid to the postorbital process suggests that the associated temporal musculature exhibited a derived configuration^{18,20,37}.

The squamosal, which to our knowledge is preserved intact in no other described Mesozoic avialan specimen between *Archaeopteryx* and *Hesperornis*, is peripheral to the braincase, unlike that of extant birds^{35,38}. The squamosal exhibits a plesiomorphic, deinonychosaur-like morphology unseen in crownward taxa: the zygomatic process widely encircles the posterior half of the upper temporal fossa, first projecting laterally and then curving anteriorly in such a way that the process is hooked and directed rostrally to form—with the postorbital process—a nearly complete upper temporal bar broken only by a small unossified and probably ligamentous gap (Fig. 4b). In lateral view, the zygomatic process is deep and triangular. As in non-avian dinosaurs, the nuchal crest along the suture between

the parietal and occiput extends from the cranial midline onto the zygomatic process, forming the upper edge of the squamosal bone (Figs. 1, 2, 4b). In its encirclement of the adductor chamber, the squamosal recalls that of much more stemward, non-ornithuran theropods (Fig. 4), in which a complete upper temporal arch is present as a retention of the ancestral diapsid condition¹⁴. The nearly complete upper temporal bar suggests plesiomorphic architecture and topology of the muscles attaching to the posterior part of the upper temporal fossa, which include the majority of the adductor externus complex³⁹.

All parts of the *Ichthyornis* skull provide insights into the form and function of the ancestral ornithuran head and the transition from an early avialan to an avian condition. The pincer-like action of a sharp-tipped, toothless beak would have facilitated fine manipulation and preening—essentially performing the role of a surrogate hand as the hands themselves became bound up into wings¹⁴. Holding and perforation of prey probably fell to the sizeable, reptilian tooth row retained in *I. dispar*²³. The concurrent appearance of a crown-grade avian kinetic apparatus would have enabled further precision in grasping and an expanded gape. Simultaneous appearance of the beak and kinetic palate is consistent with evidence for a deep molecular developmental linkage between the fusion of the premaxillary beak and the slimming and detachment of the palatines in the roof of the mouth⁴⁰; thus, the kinetic apparatus is functionally and developmentally integrated. If the kinetic apparatus was indeed produced by a discrete shift in molecular patterning, its evolutionary appearance could have been fairly abrupt or saltational. Whereas the form of the beak supports one embryologically derived hypothesis, the coexistence of a primitive adductor chamber and a derived, crown-bird-like brain challenges another: the previous suggestion that brain

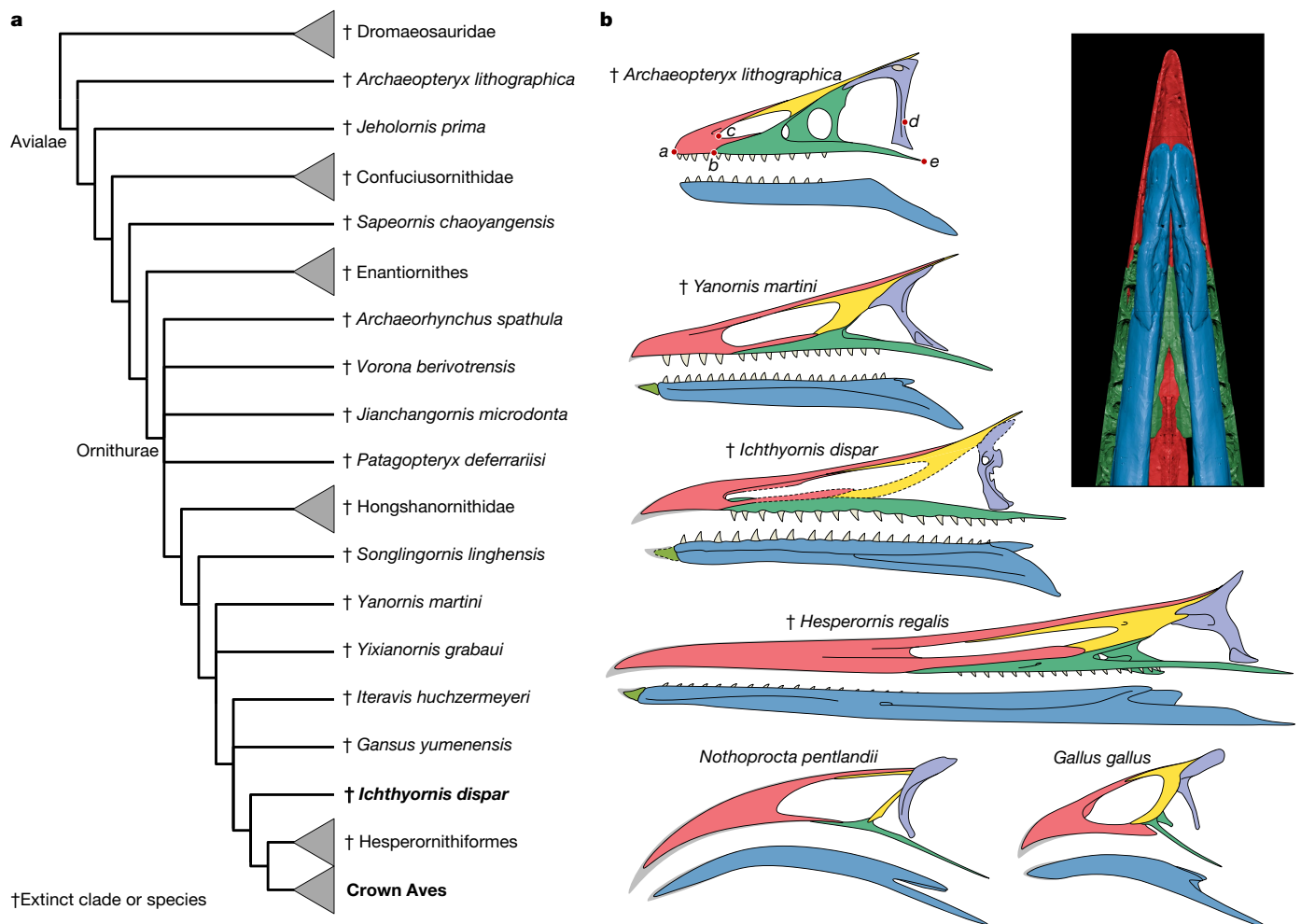


Fig. 3 | Relationships of *I. dispar* and the origin of the avian beak. **a**, Cladogram showing the phylogenetic position of *I. dispar* inferred on the basis of our analyses (abridged and based in part on new codings in multiple datasets; see Supplementary Information 1 for full phylogenetic results). **b**, Origin of the avian beak. Line drawings of the rostrum of key Mesozoic avialans and crown birds (see **a** for phylogenetic position).

Hypothesized extent of rhamphotheca is indicated in grey. Points a–e on the rostrum of *A. lithographica* denote landmarks used for comparative measurements presented in Supplementary Information 1. Inset shows the occlusal gap between the tips of the premaxillae (red) and dentaries (blue) in ventral view, inferred to have been occupied by a predentary.

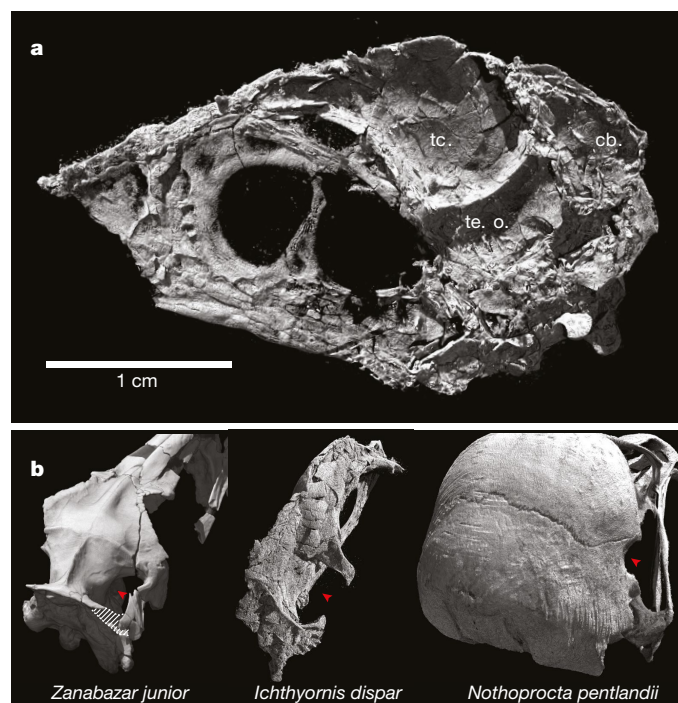


Fig. 4 | Derived brain shape and primitive temporal region of *I. dispar*.

a, Sagittal cutaway of the braincase of *I. dispar* revealing the endocranial space. Cavities labelled for brain divisions as follows: tc., telencephalon (forebrain); te. o., tectum opticum (optic lobe); cb., cerebellum.

b, Comparative views of the temporal region of a nonavian dinosaur (*Zanabazar junior*), *I. dispar* and a crown bird (Andean tinamou, *Nothoprocta pentlandii*). Red arrows indicate medial embayment of the upper temporal fenestra.

enlargement drove adductor reduction owing to spatial restriction during embryonic development¹⁴. Finally, the essentially avian brain of *I. dispar* coincides with a flight apparatus resembling that of strong fliers among living birds—evolutionary elaboration of the bird brain may have been in service to the exigencies of avian flight, the most sophisticated and demanding form of locomotion in the history of vertebrate life³⁴.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0053-y>.

Received: 5 October 2017; Accepted: 21 March 2018;

Published online 2 May 2018.

- Huang, J. et al. A new ornithurine from the Early Cretaceous of China sheds light on the evolution of early ecological and cranial diversity in birds. *PeerJ* **4**, e1765 (2016).
- Marsh, O. C. Notice of a new and remarkable fossil bird. *Am. J. Sci.* **4** (third series), 344 (1872).
- Clarke, J. A. Morphology, phylogenetic taxonomy, and systematics of *Ichthyornis* and *Apatornis* (Avialae: Ornithurae). *Bull. Am. Mus. Nat. Hist.* **286**, 1–179 (2004).
- Olson, S. L. *Ichthyornis* in the Cretaceous of Alabama. *Wilson Bull.* **87**, 103–105 (1975).
- Lucas, S. G. & Sullivan, R. M. *Ichthyornis* in the Late Cretaceous Mancos Shale (Juana Lopez Member), northwestern New Mexico. *J. Paleontol.* **56**, 545–547 (1982).
- Parris, D. & Echols, J. The fossil bird *Ichthyornis* in the Cretaceous of Texas. *Tex. J. Sci.* **44**, 201–212 (1992).
- Shimada, T. R. & Wilson, L. E. A new specimen of the Late Cretaceous bird, cf. *Ichthyornis* sp., from the Cenomanian of central Kansas, with comments on the size distribution of *Ichthyornis* in North America. *Trans. Kans. Acad. Sci.* **119**, 231–237 (2016).
- Gingerich, P. D. A new partial mandible of *Ichthyornis*. *Condor* **74**, 471–473 (1972).
- Chiappe, L. M. & Qingjin, M. *Birds of Stone: Chinese Avian Fossils from the Age of Dinosaurs* (Johns Hopkins Univ. Press, Baltimore, 2016).
- Felice, R. N. & Goswami, A. Developmental origins of mosaic evolution in the avian cranium. *Proc. Natl Acad. Sci. USA* **115**, 555–560 (2018).
- Mayr, G. *Avian Evolution: the Fossil Record of Birds and its Paleobiological Significance* (John Wiley & Sons, Oxford, 2016).
- Gauthier, J. in *The Origin of Birds and the Evolution of Flight (Memoirs of the California Academy of Sciences 8)* (ed. Padian, K.) 1–55 (California Academy of Sciences, San Francisco, 1986).
- Zhou, Z. & Martin, L. D. Distribution of the predentary bone in Mesozoic ornithurine birds. *J. Syst. Palaeontology* **9**, 25–31 (2011).
- Bhullar, B. A. S. et al. How to make a bird skull: major transitions in the evolution of the avian cranium, pedomorphosis, and the beak as a surrogate hand. *Integr. Comp. Biol.* **56**, 389–403 (2016).
- O'Connor, J. The trophic habits of early birds. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* <https://doi.org/10.1016/j.palaeo.2018.03.006> (2018).
- Zheng, X., O'Connor, J. K., Wang, X., Wang, Y. & Zhou, Z. Reinterpretation of a previously described Jehol bird clarifies early trophic evolution in the Ornithuromorpha. *Proc. R. Soc. B* **285**, 20172494 (2018).
- Hieronymus, T. L. & Witmer, L. M. Homology and evolution of avian compound ramphothecae. *Auk* **127**, 590–604 (2010).
- Baumel, J. J. et al. *Handbook of Avian Anatomy: Nomina Anatomica Avium (Publications of the Nuttall Ornithological Club, No. 23)* 2nd edn (Harvard Univ. Nuttall Ornithological, Cambridge, 1993).
- Elzanowski, A. New observations on the skull of *Hesperornis* with reconstructions of the bony palate and otic region. *Postilla* **207**, 1–20 (1991).
- Bock, W. J. Kinetics of the avian skull. *J. Morphol.* **114**, 1–42 (1964).
- Homberger, D. G. in *The Biology of the Avian Respiratory System* (ed. Maina, J. N.) 27–97 (Springer, Cham, 2017).
- Martin, L. D. & Stewart, J. D. Teeth in *Ichthyornis* (Class: Aves). *Science* **195**, 1331–1332 (1977).
- Dumont, M. et al. Synchrotron imaging of dentition provides insights into the biology of *Hesperornis* and *Ichthyornis*, the “last” toothed birds. *BMC Evol. Biol.* **16**, 178 (2016).
- Jollie, M. T. The head skeleton of the chicken and remarks on the anatomy of this region in other birds. *J. Morphol.* **100**, 389–436 (1957).
- Zusi, R. L. in *The Skull, Volume 2: Patterns of Structural and Systematic Diversity* (eds Hanken, J. & Hall, B. K.) Ch. 8, 391–437 (Univ. of Chicago Press, Chicago, 1993).
- Chiappe, L. M., Norell, M. & Clark, J. A new skull of *Gobipteryx minuta* (Aves: Enantiornithes) from the Cretaceous of the Gobi Desert. *Am. Mus. Novit.* **3346**, 1–15 (2001).
- Bühler, P., Martin, L. D. & Witmer, L. M. Cranial kinesis in the Late Cretaceous birds *Hesperornis* and *Parahesperornis*. *Auk* **105**, 111–122 (1988).
- Zusi, R. L. A functional and evolutionary analysis of rhynchokinesis in birds. *Smithson. Contrib. Zool.* **395**, 1–40 (1984).
- Gussekloof, S. W. & Bout, R. G. Cranial kinesis in palaeognathous birds. *J. Exp. Biol.* **208**, 3409–3419 (2005).
- Rauhut, O. W. M. New observations on the skull of *Archaeopteryx*. *Palaontol. Z.* **88**, 211–221 (2014).
- Mayr, G., De Pietri, V. L., Scofield, R. P. & Worthy, T. H. On the taxonomic composition and phylogenetic affinities of the recently proposed clade Vegaviidae Agnolín et al., 2017—neornithine birds from the Upper Cretaceous of the Southern Hemisphere. *Cretac. Res.* **86**, 178–185 (2018).
- Walsh, S. A., Milner, A. C. & Bourdon, E. A reappraisal of *Cerebavis cenomanica* (Aves, Ornithurae), from Melovodka, Russia. *J. Anat.* **229**, 215–227 (2016).
- Alonso, P. D., Milner, A. C., Ketcham, R. A., Cookson, M. J. & Rowe, T. B. The avian nature of the brain and inner ear of *Archaeopteryx*. *Nature* **430**, 666–669 (2004).
- Balanoff, A. M., Bever, G. S., Rowe, T. B. & Norell, M. A. Evolutionary origins of the avian brain. *Nature* **501**, 93–96 (2013).
- Elzanowski, A. & Galton, P. M. Braincase of *Enaliornis*, an Early Cretaceous bird from England. *J. Vertebr. Paleontol.* **11**, 90–107 (1991).
- Elzanowski, A. & Mayr, G. Multiple origins of secondary temporal fenestrae and orbitozygomatic junctions in birds. *J. Zool. Syst. Evol. Res.* <https://doi.org/10.1111/jzs.12196> (2017).
- Müller, H. J. Die morphologie und entwicklung des craniums von *Rhea americana* Linné. II. Viszeralskelett, mittelohr und osteocranium. *Z. Wiss. Zool.* **168**, 35–118 (1963).
- Cracraft, J. The origin and early diversification of birds. *Paleobiology* **12**, 383–399 (1986).
- Holliday, C. M. & Witmer, L. M. Archosaur adductor chamber evolution: integration of musculoskeletal and topological criteria in jaw muscle homology. *J. Morphol.* **268**, 457–484 (2007).
- Bhullar, B. A. S. et al. A molecular mechanism for the origin of a key evolutionary innovation, the bird beak and palate, revealed by an integrative approach to major transitions in vertebrate history. *Evolution* **69**, 1665–1677 (2015).

Acknowledgements We thank M. Colbert, J. Maisano and the staff of the UTCT facility at The University of Texas at Austin, as well as G. Lin at the Harvard Center for Nanoscale Systems and A. Pritchard at Yale for CT scanning assistance. K. Zyskowski and G. Watkins-Colwell in the Division of Vertebrate Zoology at YPM assisted with extant comparative material. M. Fox, C. Norris and D. Brinkman facilitated the examination and scanning of YPM fossil material. This research was supported by Yale University, the Yale Peabody Museum of Natural History, the University of Bath, National Science Foundation Doctoral Dissertation Improvement Grant DEB 1500798, the Alexander Wetmore Memorial Research Award (American Ornithologists' Union), a Yale Institute for Biospheric Studies Dissertation Improvement Grant, the Stephen J. Gould Award (Paleontological Society), an Evolving Earth Foundation Research Grant and a Frank M. Chapman Ornithological Research Grant (American Museum of Natural History).

Reviewer information Nature thanks Z. Zhou and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.J.F. and B.-A.S.B. conceived and directed the study and arranged logistics of specimen preparation and CT scanning. K.S. discovered the FHSM specimen and donated it to the museum. D.J.F., M.H. and B.-A.S.B. performed CT scans and processed CT data and M.H. assembled the skull reconstruction. M.H., D.J.F. and B.-A.S.B. scored characters and performed phylogenetic analyses. B.-A.S.B. and M.H. planned the main-text figures. M.H. and D.J.F. prepared the figures. D.J.F. wrote the supplementary anatomical descriptions and rendered the Supplementary Videos. D.B., L.E.W., K.S., D.E. and J.A.E. collected and prepared specimens for study, analysed morphology and edited the paper. B.-A.S.B., D.J.F. and M.H. wrote the paper. B.-A.S.B. and D.J.F. acquired funding.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0053-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0053-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to B.-A.S.B. **Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

I. dispar specimens housed at ALMNH, BHI, FHSM, KUVV and YPM were scanned at the University of Texas High-Resolution CT Facility (UTCT) and the Center for Nanoscale Systems at Harvard. Scan parameters and specimen details are presented in Supplementary Information 1.

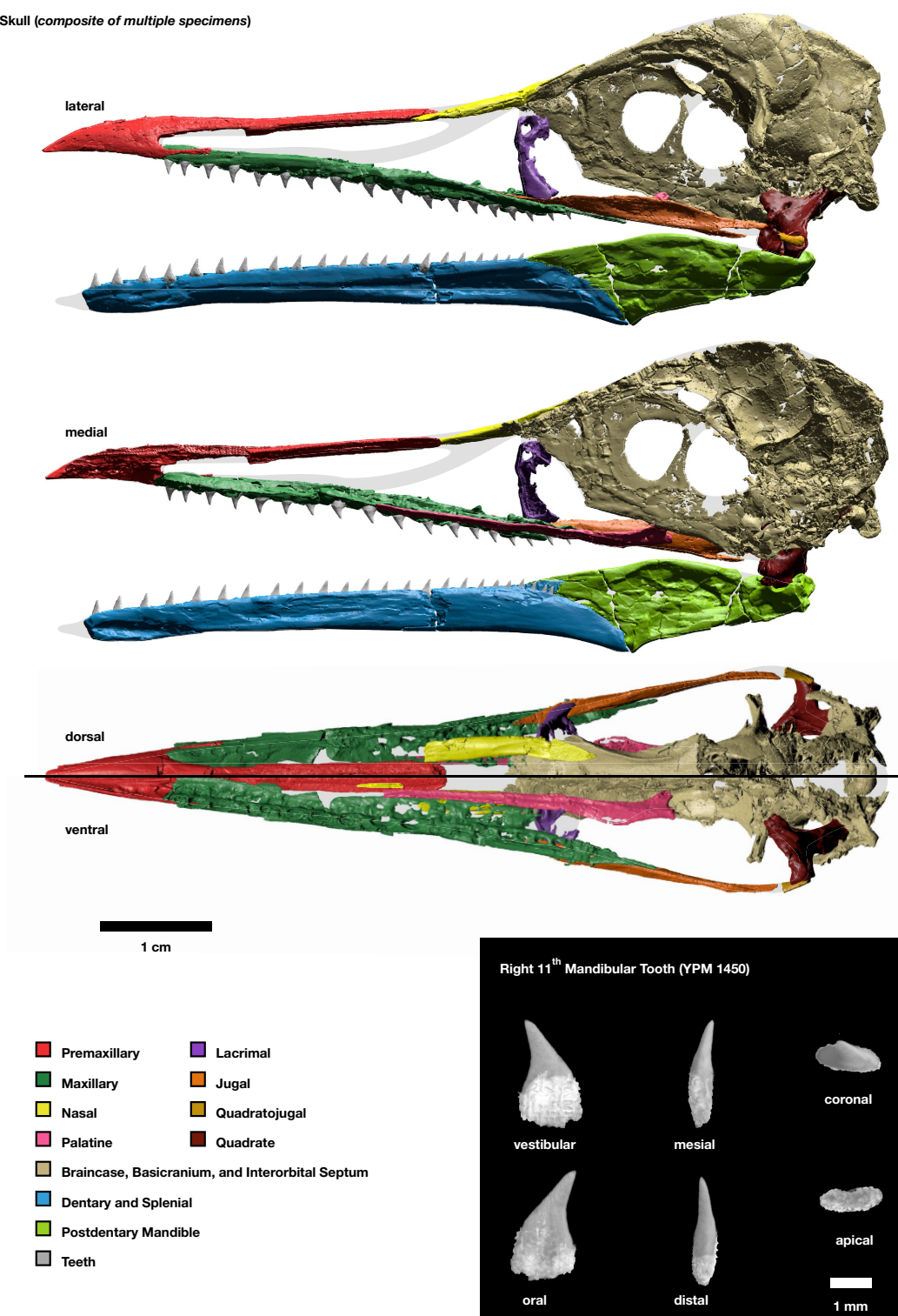
Scanned cranial material was digitally segmented using VGStudio MAX 3.0, and 3D surface meshes were extracted and imported to MeshLab 2016 for optimization. Optimized meshes were then assembled into 3D reconstructions using Autodesk Maya 2017.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

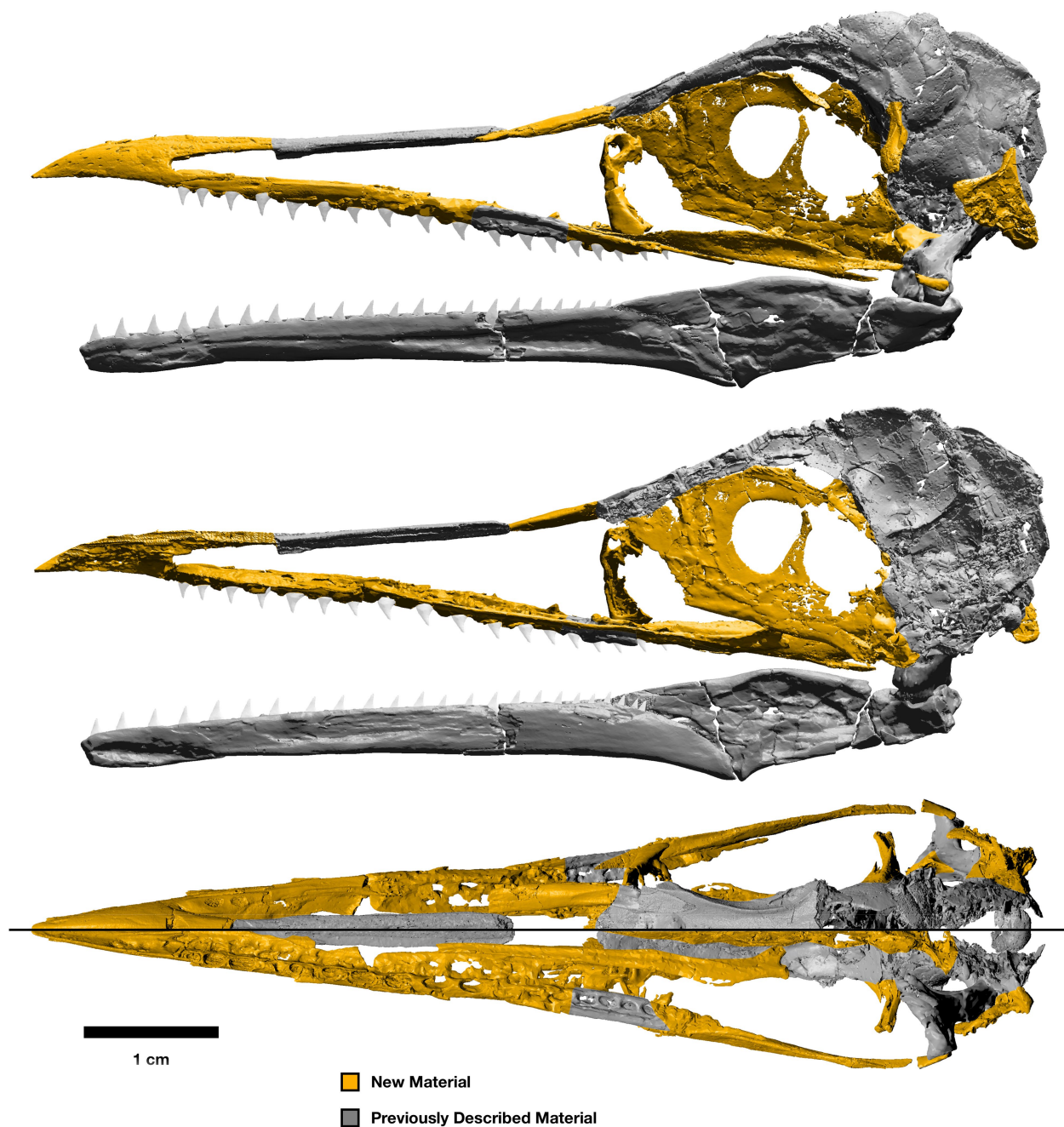
Data availability. Phylogenetic matrices and detailed images of anatomy are included as Supplementary Information. Three-dimensional models and data are archived and available on request from the Yale Peabody Museum of Natural History. All other data are available from the corresponding author upon reasonable request.

Ichthyornis dispar (Marsh, 1872)

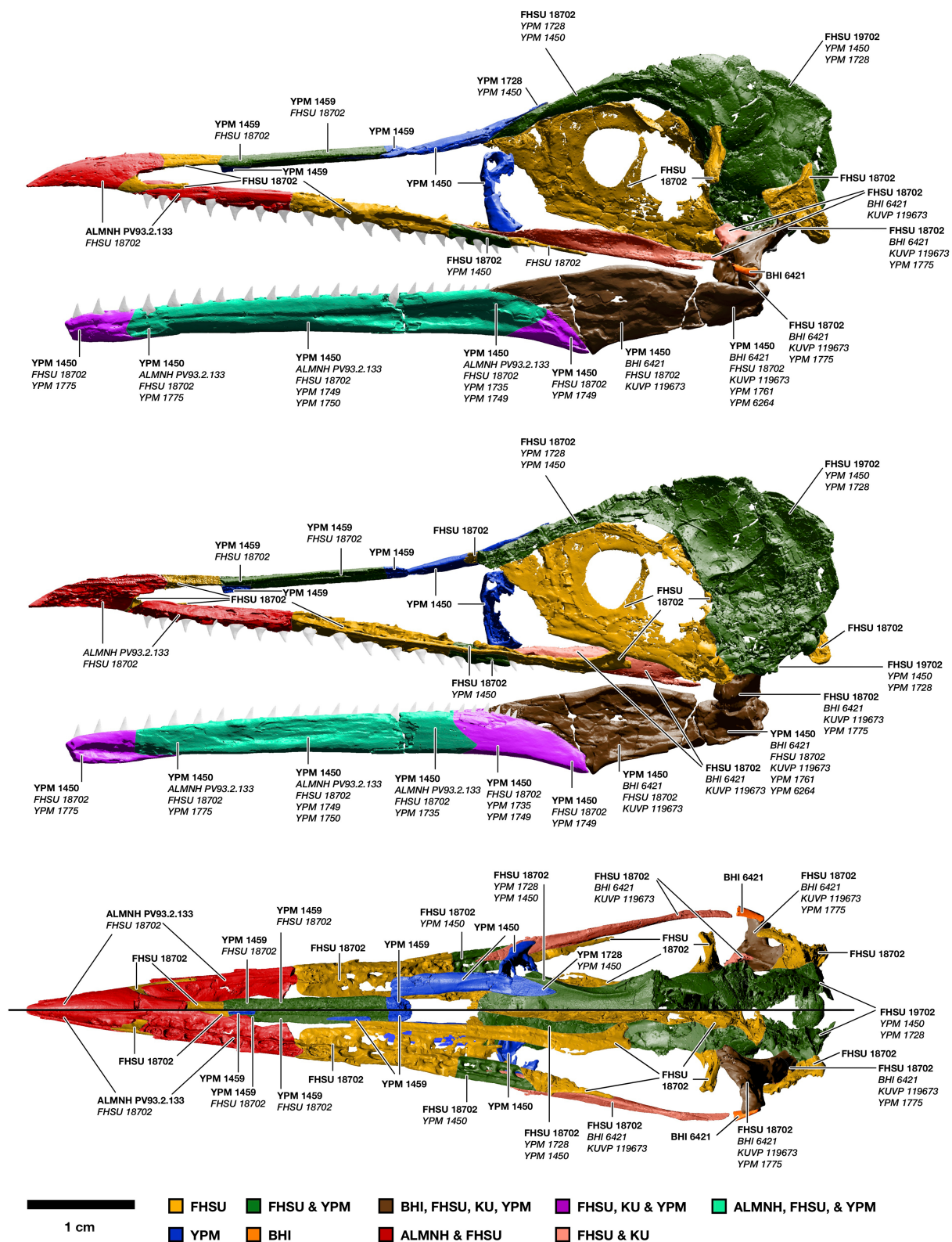
Skull (composite of multiple specimens)



Extended Data Fig. 1 | Full 3D reconstruction of the skull of *I. dispar* in high resolution. This is the same reconstruction as shown in Fig. 1, reproduced at a higher resolution to show details.



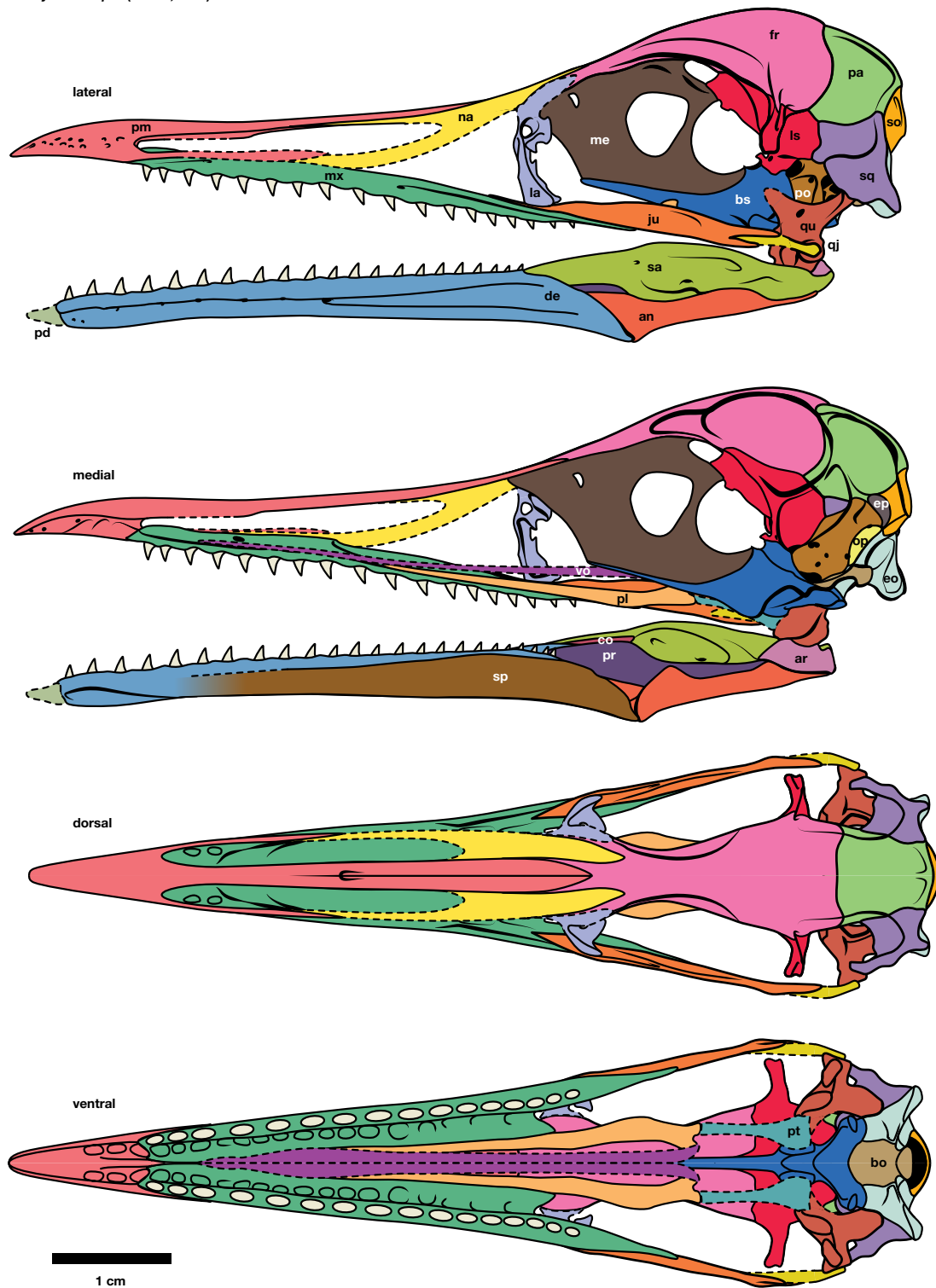
Extended Data Fig. 2 | Reconstruction of the skull of *I. dispar*. Material described in this paper is indicated in gold and previously described regions are indicated in grey. All elements are scaled to the size of the FHSM VP-18702 specimen.



Extended Data Fig. 3 | Reconstruction of the skull of *I. dispar* indicating the material represented by every known *Ichthyornis* specimen. All elements are scaled to the size of the FHSU VP-18702

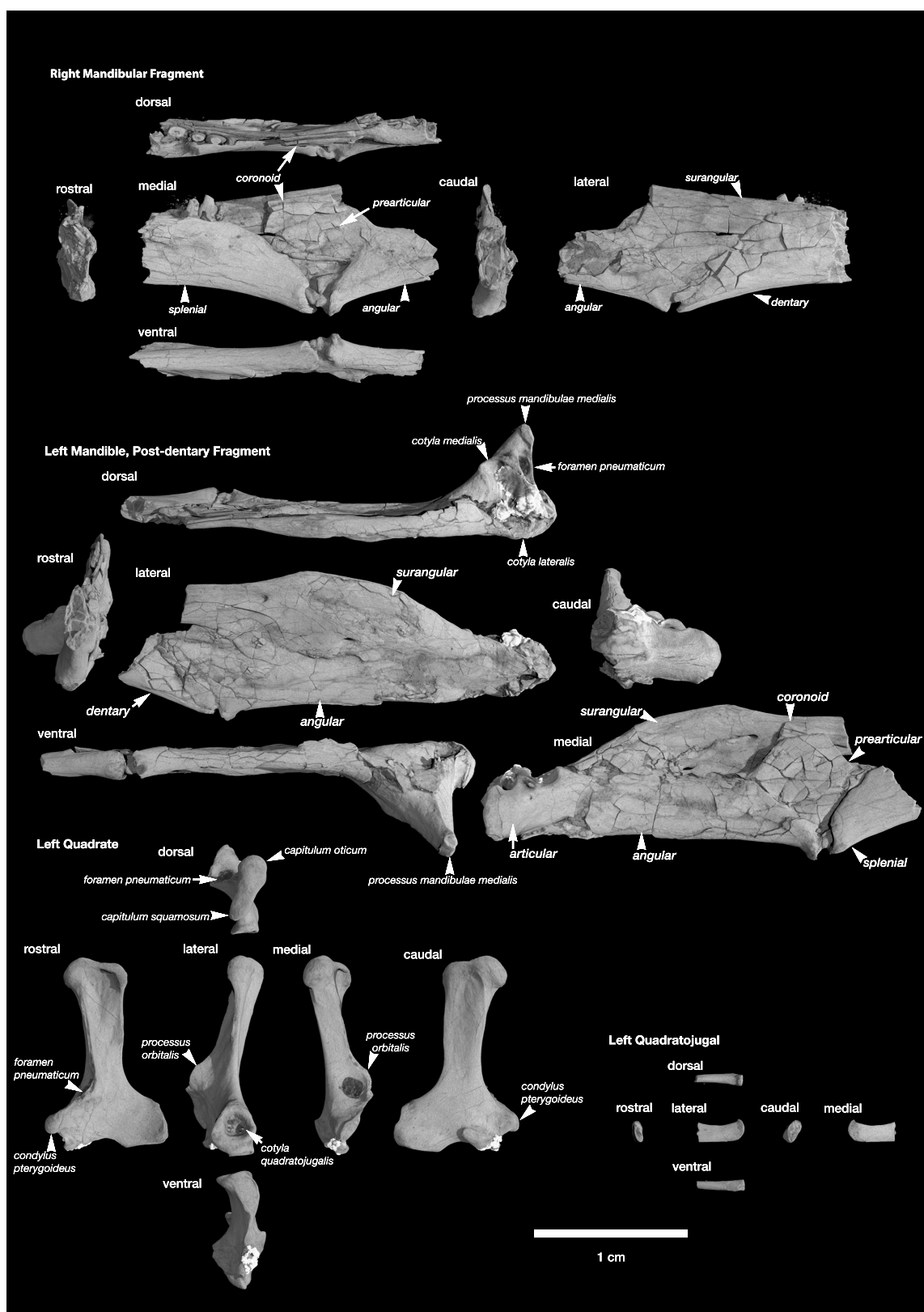
specimen. Specimen numbers in bold are those used in the reconstruction. Numbers in italics indicate preservation of the same element in additional specimens.

Ichthyornis dispar (Marsh, 1872)

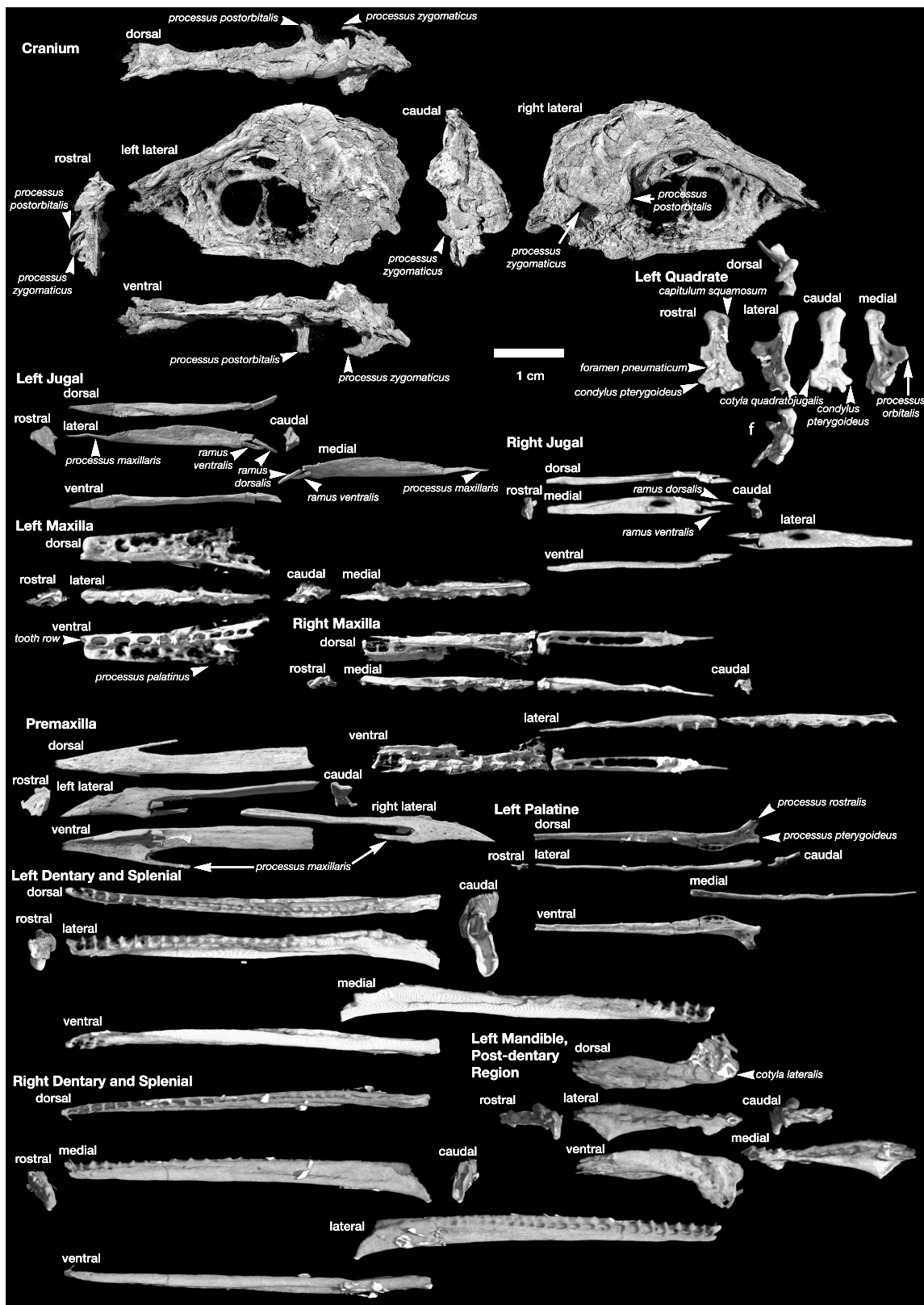


Extended Data Fig. 4 | High-resolution line drawing of the skull of *I. dispar*. This the same image as shown in Fig. 2, reproduced at a larger size to show details. All anatomical abbreviations are as indicated in

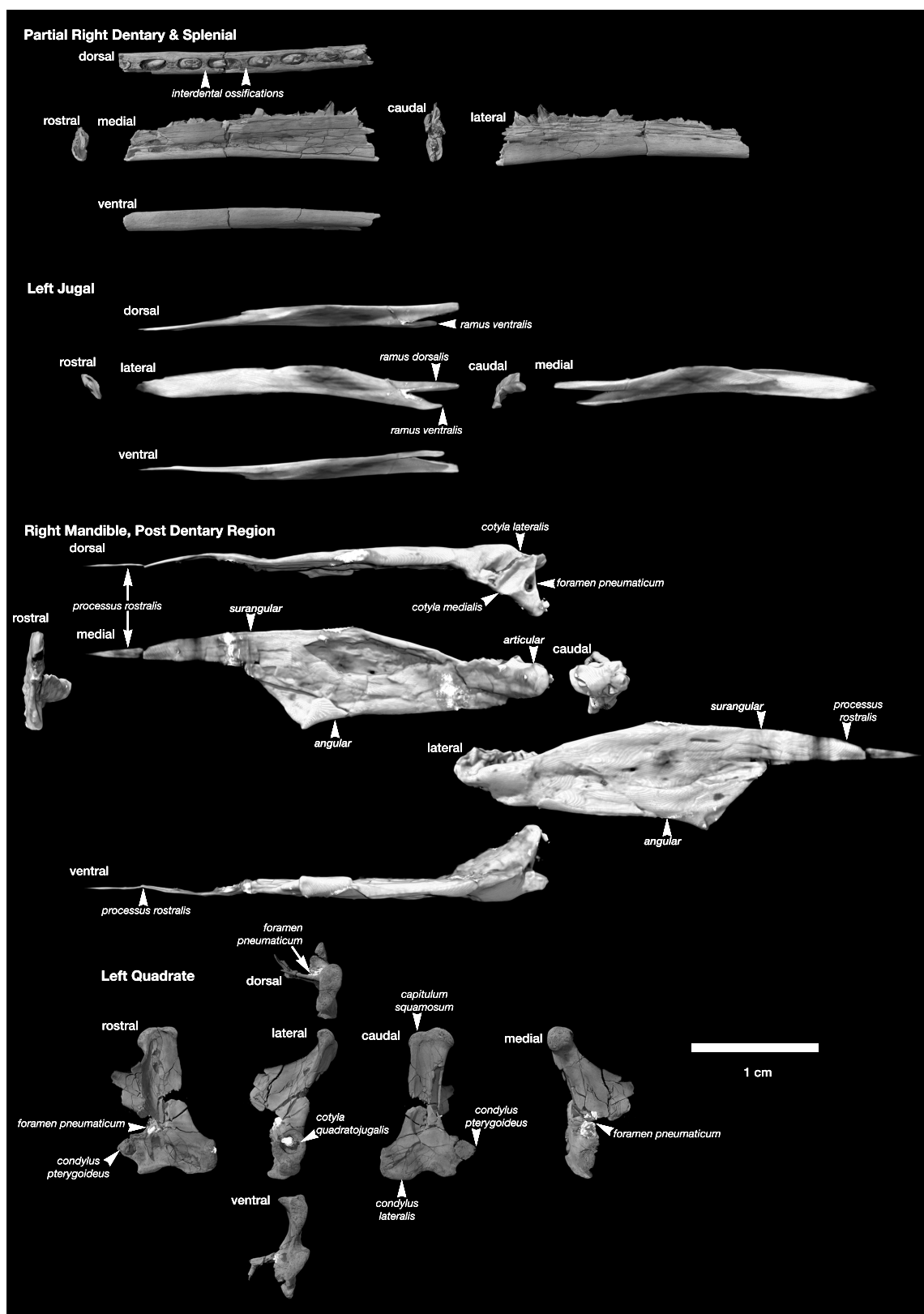
Fig. 2. Solid lines indicate areas known from fossil specimens, and dashed lines indicate unknown areas reconstructed from other ornithuran birds.



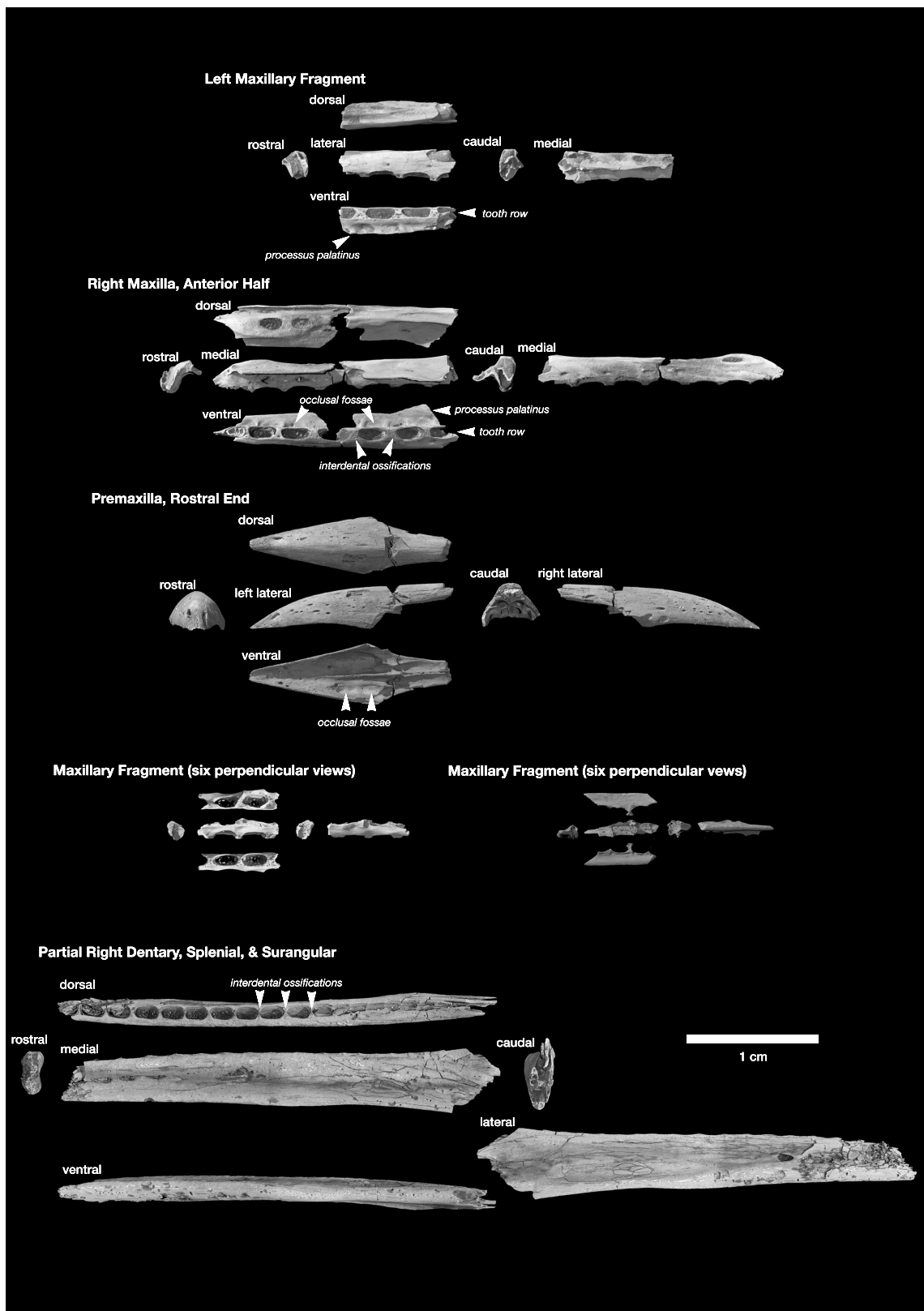
Extended Data Fig. 5 | Skull and jaw elements of *I. dispar* specimen BHI 6421.



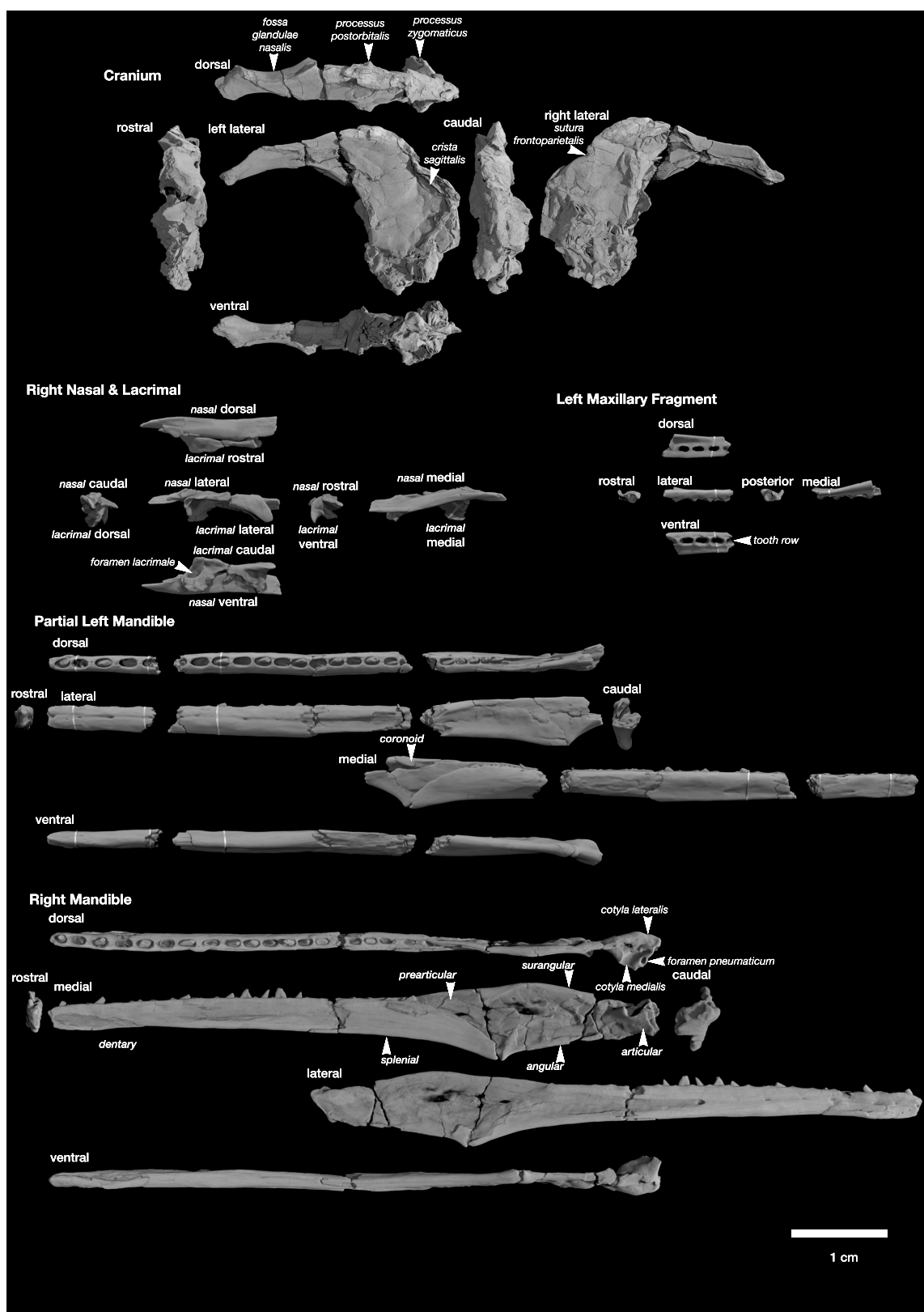
Extended Data Fig. 6 | Skull and jaw elements of *I. dispar* specimen FHSM VP-18702.



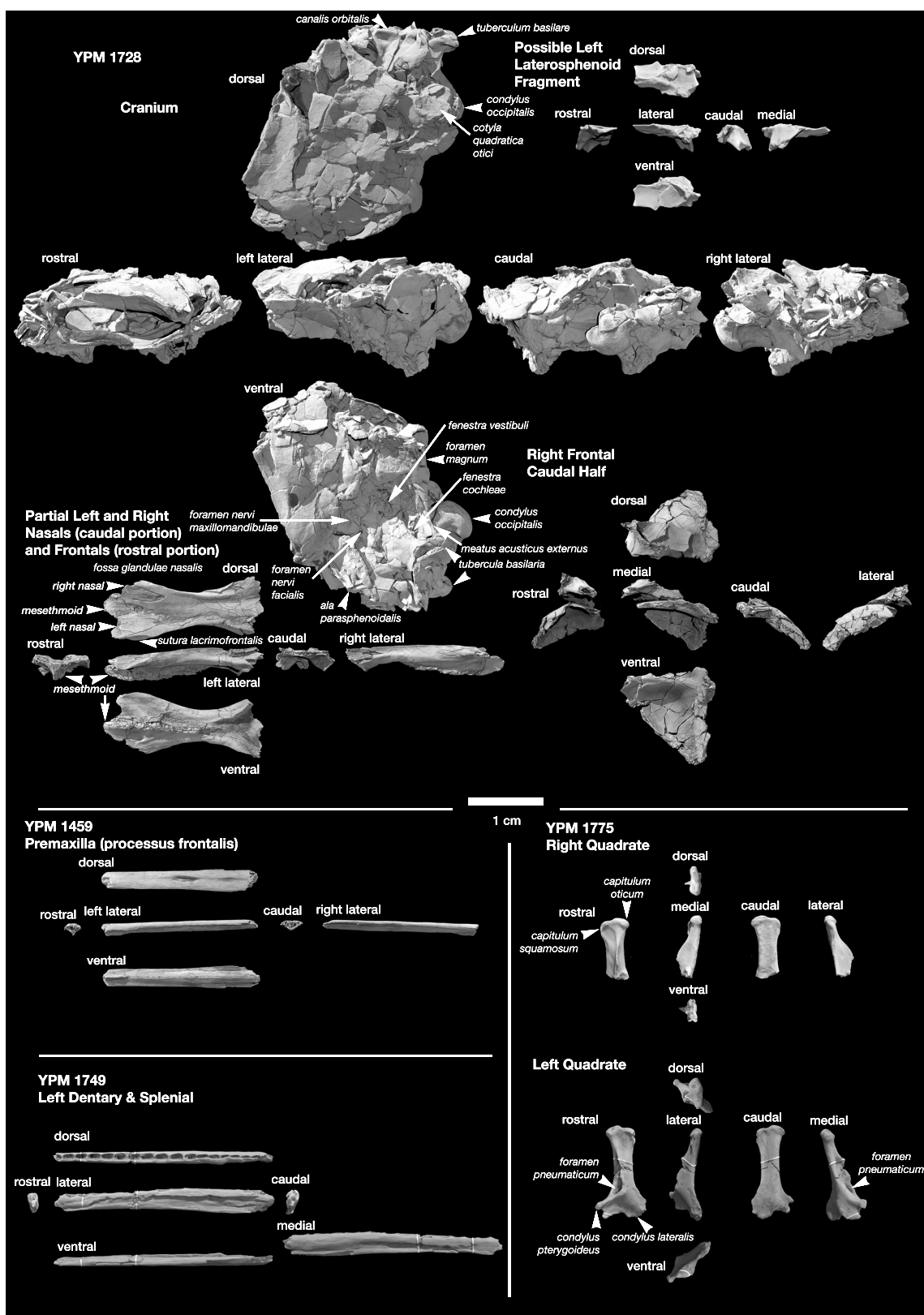
Extended Data Fig. 7 | Skull and jaw elements of *I. dispar* specimen KUV 119673.



Extended Data Fig. 8 | Skull and jaw elements of *I. dispar* specimen ALMNH 3316.



Extended Data Fig. 9 | Skull and jaw elements of *I. dispar* holotype YPM 1450 showing the nasal and lacrimal elements that have not previously been reported.



Extended Data Fig. 10 | Skull and jaw elements of *I. dispar* specimens YPM 1728, YPM 1459, YPM 1775 and YPM 1749.

Deep mitochondrial origin outside the sampled alphaproteobacteria

Joran Martijn¹, Julian Vosseberg^{1,5}, Lionel Guy², Pierre Offre^{3,4} & Thijs J. G. Ettema^{1*}

Mitochondria are ATP-generating organelles, the endosymbiotic origin of which was a key event in the evolution of eukaryotic cells¹. Despite strong phylogenetic evidence that mitochondria had an alphaproteobacterial ancestry², efforts to pinpoint their closest relatives among sampled alphaproteobacteria have generated conflicting results, complicating detailed inferences about the identity and nature of the mitochondrial ancestor. While most studies support the idea that mitochondria evolved from an ancestor related to Rickettsiales^{3–9}, an order that includes several host-associated pathogenic and endosymbiotic lineages^{10,11}, others have suggested that mitochondria evolved from a free-living group^{12–14}. Here we re-evaluate the phylogenetic placement of mitochondria. We used genome-resolved binning of oceanic metagenome datasets and increased the genomic sampling of Alphaproteobacteria with twelve divergent clades, and one clade representing a sister group to all Alphaproteobacteria. Subsequent phylogenomic analyses that specifically address long branch attraction and compositional bias artefacts suggest that mitochondria did not evolve from Rickettsiales or any other currently recognized alphaproteobacterial lineage. Rather, our analyses indicate that mitochondria evolved from a proteobacterial lineage that branched off before the divergence of all sampled alphaproteobacteria. In light of this new result, previous hypotheses on the nature of the mitochondrial ancestor^{6,15,16} should be re-evaluated.

Mitochondria are important organelles in eukaryotic cells and are involved in various processes, of which ATP generation through oxidative phosphorylation is a hallmark feature. The endosymbiotic origin of these organelles represents a strongly debated step in eukaryotic evolution¹⁷ that has contributed to the emergence of the cellular complexity that characterizes modern eukaryotes¹. To trace the evolutionary history of mitochondria and their role in eukaryogenesis, detailed knowledge about the identity and nature of the mitochondrial ancestor is of great importance. However, despite the fact that the alphaproteobacterial origin of mitochondria is generally undisputed, efforts to resolve the phylogenetic position of mitochondria in the alphaproteobacterial species tree have failed to reach consensus owing to several complications. First, because mitochondria and several independent alphaproteobacterial lineages have been subjected to accelerated rates of evolution, they are sensitive to long branch attraction (LBA) artefacts in which fast evolving lineages incorrectly group together^{3,8,18}. Second, because alphaproteobacterial and mitochondrial sequences display a high degree of compositional heterogeneity, they are sensitive to the compositional bias artefact in which unrelated lineages with similar sequence compositions falsely group together^{3,8,19,20}. Finally, the current sample of alphaproteobacterial genomes is biased towards taxa that are clinically or agriculturally relevant or can be cultivated in a laboratory setting. It does not reflect the natural diversity of extant alphaproteobacteria and might exclude potential close relatives of mitochondria. Here we attempt to resolve the phylogenetic origin of mitochondria by addressing the issues of long branch attraction, compositional bias

and biased taxon sampling simultaneously. We tried to remove these phylogenetic artefacts by applying models of sequence evolution that account for site-specific substitution patterns, by reducing compositional heterogeneity of molecular sequence data and by increasing taxon sampling in a more unbiased manner through genome-resolved metagenomic binning.

We screened all publicly available *Tara* Oceans metagenomic datasets²¹ for novel alphaproteobacteria residing in the ocean's upper layers (see Methods) and selected three datasets corresponding to samples taken at 5 m, 115 m and 140 m depths in the Pacific Ocean (Fig. 1a). To also capture novel alphaproteobacteria residing in the deeper layers, we screened four metagenomic datasets²² originating from samples taken at 100 m, 776 m, 2,745 m and 5,002 m depths in the Atlantic Ocean (Fig. 1a).

We assembled each dataset with recently developed metagenome assemblers^{23,24} and performed a phylogenomic analysis of all contigs that contained at least five out of fifteen ribosomal proteins located in the str-spc cluster²⁵ (RP15). Collectively, the seven metagenomes contained a large diversity of alphaproteobacteria (Supplementary Fig. 1). In particular, they harboured several novel divergent lineages potentially useful for resolving the phylogenetic position of mitochondria (Fig. 1b and Supplementary Fig. 2).

We reconstructed genomes from these lineages by binning metagenomic contigs into metagenome-assembled genomes (MAGs) based on their differential sequence coverage across samples, tetranucleotide frequencies and read-pair linkage information. Forty-five MAGs were reconstructed that exhibited high completeness and low redundancy estimates, collectively representing twelve distinct alphaproteobacterial lineages (designated MarineAlpha1–12; Fig. 2 and Supplementary Table 1) and one additional proteobacterial lineage that forms a sister clade to the Alphaproteobacteria (designated MarineProteo1; Fig. 2). The genome sample of the obtained MAGs is extensive, as they covered most divergent groups observed in the RP15 phylogeny (Fig. 1b and Supplementary Fig. 2). None of the MAGs contained full-length 16S rRNA genes, a commonly observed issue²⁶. This prevented us from linking these lineages to alphaproteobacteria that have been previously identified in 16S rRNA environmental surveys. The exception was MarineAlpha12, which lacks close relatives in the SILVA database (release 128; <https://www.arb-silva.de/>) to give it a meaningful taxonomic classification ('uncultured Rhodospirillaceae' at 92.9% similarity).

Next, we determined the phylogenetic positions of the MAGs. We performed phylogenomic analyses on a concatenated dataset comprising 72 carefully selected genes that are conserved across Alphaproteobacteria (Supplementary Table 2). Bayesian and maximum-likelihood phylogenetic inferences of this dataset generally yielded species trees in which all AT-rich (<40% GC), long branch taxa (Rickettsiales, Pelagibacteraceae, alphaproteobacterium HIMB59, MarineAlpha2, 5–8 and MarineAlpha9 bins 1–4) formed a monophyletic group with maximum statistical support (Supplementary Figs. 3, 4).

¹Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ²Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ³Max Planck Institute for Marine Microbiology, Bremen, Germany. ⁴Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research and Utrecht University, Den Burg, The Netherlands. ⁵Present address: Department of Biology, Utrecht University, Utrecht, The Netherlands. *e-mail: thijs.ettema@icm.uu.se

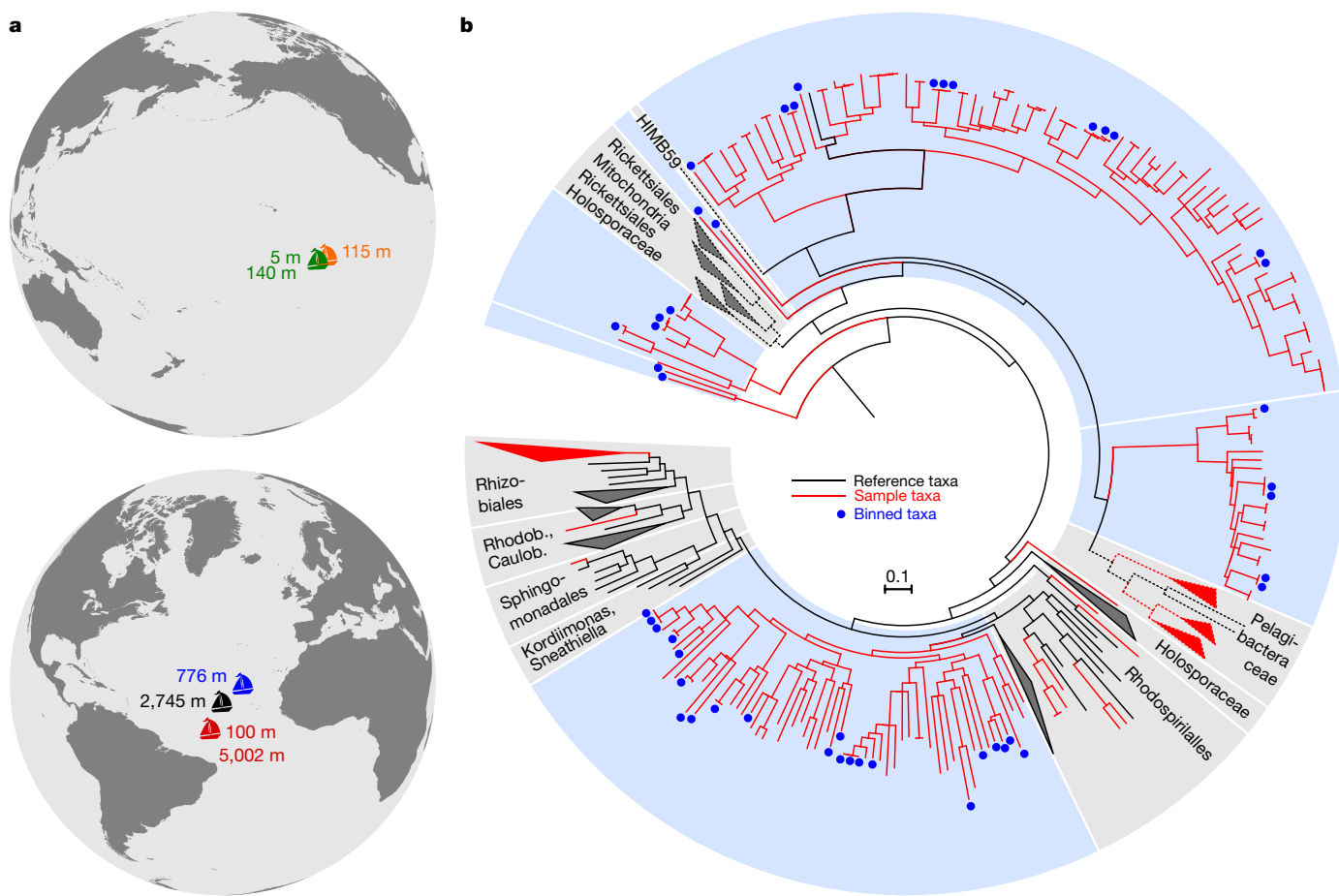


Fig. 1 | Metagenomic exploration of oceanic alphaproteobacteria. Locations, depths and alphaproteobacterial phylogenetic diversity of the seven environmental samples. **a**, Sample depths and approximate locations. Coordinates (latitude, longitude) of the Atlantic Ocean samples were as follows: 100 m and 5,002 m: 10.91°, 315.33° (red); 776 m: 24.51°, 325.8° (blue); 2,745 m: 20.12°, 319.16° (black). Coordinates of the Pacific

Ocean samples can be found in Supplementary Table 6. **b**, Phylogenetic diversity of sampled alphaproteobacteria compared to a reference set of alphaproteobacteria and mitochondria based on an RP15 analysis (see Methods). Branches that have been shortened for clarity purposes are dotted. Caulob., Caulobacteriales; Rhodob., Rhodobacteriales. For a fully expanded tree see Supplementary Fig. 2.

Such a topology strongly suggests that, despite the use of mixture models of evolution, these analyses suffered from LBA and compositional bias artefacts. Indeed, posterior predictive tests of the Bayesian analysis confirmed that the used model of evolution failed to adequately capture the level of compositional heterogeneity in the dataset (Supplementary Fig. 3 and Supplementary Table 3).

To ameliorate compositional bias artefacts, we reduced the compositional heterogeneity of the dataset by either recoding the data from a 20-character to a 4-character state²⁷, or by removing heterogeneous sites until all pairwise taxa combinations were considered homogeneous with a stationary-based trimmer²⁸. Phylogenetic analyses of the compositionally homogenized datasets appeared to be free of compositional bias artefacts: the Pelagibacteraceae, Rickettsiales, MarineAlpha9 bins 1–4 and MarineAlpha2 bin 1 each branched in distinct areas of the tree, away from alphaproteobacterium HIMB59 and MarineAlpha5–8 (Fig. 3, Supplementary Figs. 5–8 and Supplementary Discussion). Posterior predictive tests of Bayesian analyses of the compositionally homogenized datasets indicated that the evolutionary model was either substantially closer to capturing the level of compositional heterogeneity or was able to capture it adequately (Supplementary Table 3). Reducing heterogeneity alleviates artefacts, but also reduces the informational content, which in turn leads to a decreased statistical support for some of the deeper nodes of the trees (Fig. 3 and Supplementary Figs. 5–8). Notably, the phylogenetic analyses of the compositionally homogenized datasets reveal a species tree in which the MAGs represent clades that do not branch closely with any of the previously established groups. Therefore, the

addition of the MAGs captures a larger fraction of the alphaproteobacterial diversity.

Using this improved genome sample, we re-evaluated the origin of mitochondria. To do so, we assembled a phylogenomics dataset that consisted of 24 genes, which are conserved in Alphaproteobacteria and in a set of gene-rich mitochondrial genomes from diverse eukaryotes. This dataset was analysed by using the abovementioned approaches to recover phylogenies minimally affected by phylogenetic artefacts.

Phylogenetic inference from the untreated and therefore compositionally heterogeneous dataset yielded trees in which mitochondria branched together with other AT-rich and long branch taxa with high branch support (Supplementary Figs. 9, 10), suggesting that the topology was affected by compositional bias artefacts. Indeed, posterior predictive tests indicated that the model was unable to adequately capture the level of compositional heterogeneity (Supplementary Table 3). As shown for the alphaproteobacterial species tree, reducing the compositional heterogeneity reduced the number of artefactual phylogenetic relationships (Fig. 4a and Supplementary Figs. 11–13) and improved posterior predictive test results (Supplementary Table 3) at the expense of phylogenetic signal. For phylogenetic reconstructions in which the model was not able to capture heterogeneity adequately (Supplementary Table 3) (using the untreated and recoded datasets), mitochondria branched with Rickettsiales (Rickettsiales-sister) with moderate to high support (Supplementary Figs. 9, 13). By contrast, reconstructions in which the model was able to capture heterogeneity adequately (Supplementary Table 3) (using the stationary-based

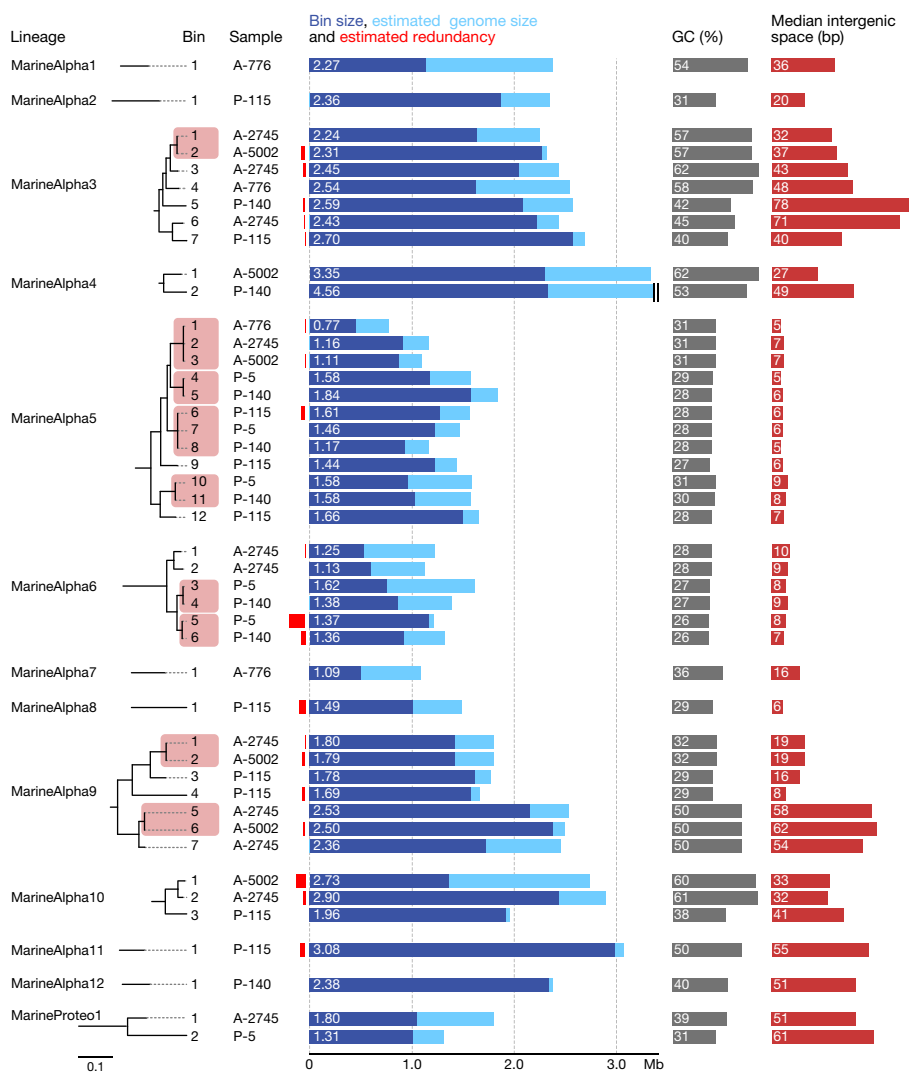


Fig. 2 | Increased genomic sampling of alphaproteobacterial clades. Overview of all genomic bins obtained in this study. The subtrees shown on the left are scaled and are based on the tree shown in Fig. 3. ‘A-’ and ‘P-’ denote Atlantic and Pacific samples, respectively. Phylogenetically

trimmed dataset), mitochondria branched as a sister group to all alphaproteobacteria (not including Magnetococcales; Supplementary Discussion) ('Alphaproteobacteria-sister') with moderate support (Fig. 4a and Supplementary Fig. 11). Similarly, approximately unbiased tests favoured a Rickettsiales-sister topology when using the untreated dataset, and favoured an Alphaproteobacteria-sister topology when using the stationary-based trimmed dataset (Supplementary Table 4). However, it should be noted that the approximately unbiased test was unable to reject the Rickettsiales-sister topology for the stationary-based trimmed dataset (Supplementary Table 4). These analyses suggest that the frequently cited Rickettsiales-sister hypothesis may be the result of a compositional attraction^{29,30} and instead point towards an Alphaproteobacteria-sister hypothesis.

The support for Alphaproteobacteria-sister in the Bayesian analysis was only moderately high (posterior probability support (PP) = 0.89). We postulated that the loss of phylogenetic information that accompanies stationary-based trimming was responsible for the lack of strong branch support. We therefore applied an alternative method (χ^2 trimmer; see Methods) that removes variable amounts of heterogeneous sites, and so allows one to retain more information. When we used the χ^2 trimmer to remove the 20% most heterogeneous sites (versus approximately 42% by the stationary-based trimmer), we recovered a similar, but generally more highly supported topology in which the support for the Alphaproteobacteria-sister was stronger (PP = 0.97; Supplementary

‘identical’ bins used to build composite bins are shaded by pink boxes. Bars indicating the estimated redundancy (red) follow the same scale as bars indicating bin size (dark blue) and estimated genome size (light blue).

Fig. 14). Although the model did not capture the heterogeneity of the χ^2 -trimmed data adequately, it was substantially closer than in the analysis of the untreated data (Supplementary Table 3). Removing fewer heterogeneous sites thus maintained more phylogenetic information and was sufficient to reduce major compositional bias artefacts. The weaker support for the Alphaproteobacteria-sister hypothesis was therefore most likely caused by the loss of phylogenetic information.

To further assess the Alphaproteobacteria-sister and Rickettsiales-sister hypotheses, we investigated the influence of compositional heterogeneity on the placement of the mitochondria by progressively reducing compositional heterogeneity of the untreated dataset with the χ^2 trimmer and inferred maximum-likelihood phylogenies. We evaluated the support for Rickettsiales-sister and Alphaproteobacteria-sister hypotheses along the heterogeneity gradient in terms of non-parametric bootstrap measures and *P* values of the approximately unbiased test. The support for Rickettsiales-sister dropped whereas support for Alphaproteobacteria-sister rose strikingly as heterogeneity decreased (Fig. 4b). In addition, we inferred maximum-likelihood phylogenies from one additional dataset that included a wider diversity of mitochondria and two additional concatenated alignments of different gene sets (Supplementary Table 2). In the first gene set, we relaxed our gene selection criterion and added 14 mitochondrial-encoded genes that displayed a lesser degree of conservation compared to the 24 originally selected genes (see Methods). The second gene set comprised

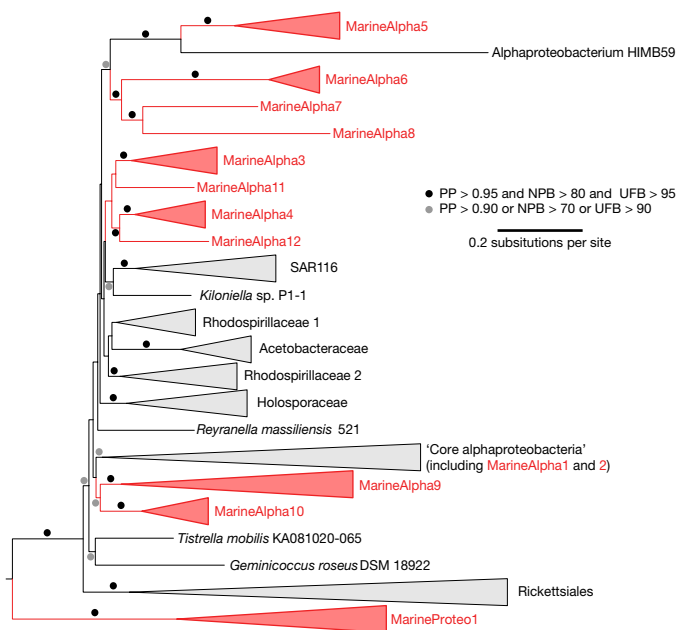


Fig. 3 | Updated alphaproteobacterial tree. Phylogenetic placement of all genomic bins relative to representatives of the genomically sampled alphaproteobacteria. Phylogenetic tree based on the stationary-trimmed alignment derived from the concatenation of 72 genes conserved across Alphaproteobacteria (Supplementary Table 2), inferred with PhyloBayes MPI under CAT + GTR + Γ 4. PP, posterior probability support; NPB, non-parametric bootstrap support; UFB, ultrafast-bootstrap support. IQTREE was used to calculate NPB under the PMSF approximation of LG + C60 + F + Γ 4 and UFB under LG + C60 + F + Γ 4. The tree is rooted with representatives of Beta-, Gammaproteobacteria and Magnetococcales (not shown). For a fully expanded tree, including taxon-to-clade specifications, see Supplementary Fig. 6.

29 nuclear-encoded genes that are thought to have been transferred from the mitochondrial genome to the nuclear genome early on in eukaryote evolution and may therefore have escaped the higher evolutionary rates and AT-biases that mitochondrial genomes have been generally subjected to. Thus, these nuclear-encoded genes are in theory less sensitive to phylogenetic artefacts than mitochondrial-encoded genes³. Maximum-likelihood trees inferred from the increased mitochondrial diversity dataset and relaxed mitochondrial-encoded gene set displayed the same behaviour as the stringent mitochondrial-encoded gene set, in which support for the Rickettsiales-sister was replaced by support for the Alphaproteobacteria-sister upon reduction of compositional heterogeneity (Supplementary Figs. 15, 16, 22, 23, Supplementary Table 4 and Supplementary Discussion). Similarly, maximum-likelihood trees inferred from the untreated nuclear-encoded gene set recovered an Alphaproteobacteria-sister topology, despite the artificial grouping of all other AT-rich taxa (Supplementary Fig. 17). In addition, approximately unbiased tests rejected Rickettsiales-sister under both LG + C60 models and its PMSF approximation (Supplementary Table 4). Maximum-likelihood trees inferred from the compositionally homogenized alignment of nuclear-encoded genes also recovered an Alphaproteobacteria-sister topology, while the AT-rich taxa no longer branched together (Supplementary Fig. 18). However, compared to maximum-likelihood trees inferred from the compositionally homogenized alignment of mitochondrial-encoded genes, branch supports for the Alphaproteobacteria-sister were lower, and approximately unbiased tests failed to discriminate between Rickettsiales- and Alphaproteobacteria-sister hypotheses (Supplementary Table 4). We hypothesize that this could be the result of a lower phylogenetic signal in the compositionally homogenized alignment of nuclear-encoded genes relative to the compositionally homogenized alignment of mitochondrial-encoded genes. The nuclear-encoded gene set had more missing data among eukaryotic taxa (30% versus 12%), displayed

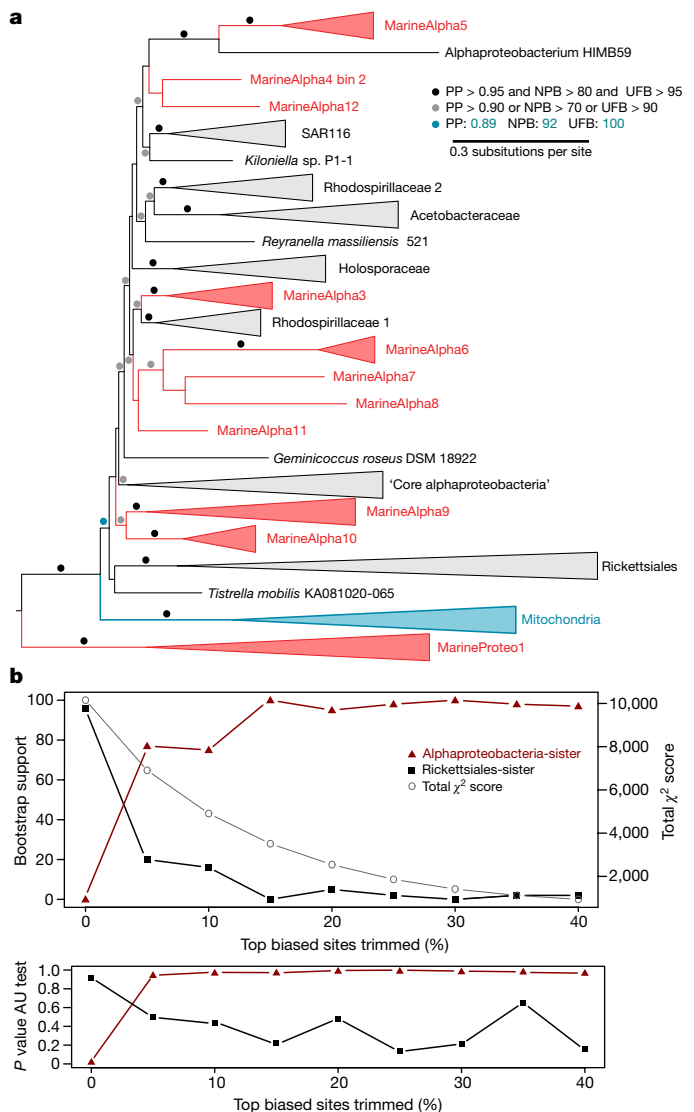


Fig. 4 | An early-branching mitochondrial ancestor. Phylogenetic placement of the mitochondria relative to representatives of the genomically sampled Alphaproteobacteria and genomic bins. **a**, Phylogenetic tree based on the stationary-trimmed alignment derived from the concatenation of 24 genes conserved across Alphaproteobacteria and gene-rich mitochondria (Supplementary Table 2), inferred with PhyloBayes MPI under CAT + GTR + Γ 4. IQTREE was used to calculate NPB under the PMSF approximation of LG + C60 + F + Γ 4 and UFB under LG + C60 + F + Γ 4. Tree is rooted with representatives of Beta-, Gammaproteobacteria and Magnetococcales (not shown). For a fully expanded tree, including taxon-to-clade specifications, see Supplementary Fig. 11. **b**, Evaluation of non-parametric bootstrap (top) and approximately unbiased (AU) test *P* value (bottom) support under the PMSF approximation of LG + C60 + F + Γ 4 for Rickettsiales-sister and Alphaproteobacteria-sister hypotheses, as sites that contribute most to overall compositional heterogeneity (total χ^2 score) are removed.

relatively higher substitution rates among eukaryotic taxa (average root-to-tip substitution rate: 1.280 versus 1.083), and probably carries a less coherent signal because it consists of more biochemically independent genes (Supplementary Table 2).

The Alphaproteobacteria-sister topology may be the result of an LBA between mitochondria and outgroup taxa. Indeed, our strategies do not directly accommodate heterotachy, and can lead to LBA. We carried out three independent analyses (outgroup removal, parametric simulations and random sequence replacements) and conclude that outgroup taxa are unlikely to have artificially attracted the mitochondria (Supplementary Discussion).

In summary, we have obtained genome data for thirteen alphaproteobacteria-related clades and re-evaluated the phylogenetic origin of the mitochondria in the context of this expanded coverage of alphaproteobacterial diversity. While none of the newly identified lineages branched specifically with mitochondria, we observed that the improved taxonomic sampling in combination with methods that address long branch attraction and compositional bias artefacts resulted in a new phylogenetic position for mitochondria. Importantly, we show that a Rickettsiales-sister origin, as reported in previous studies, is most likely the result of a compositional bias artefact. Our analyses instead suggest that mitochondria diverged from Alphaproteobacteria before the diversification of all currently known alphaproteobacterial lineages. We infer that Rickettsiales and mitochondria evolved via two independent endosymbiosis events, instead of the single shared endosymbiosis event¹⁵ that is suggested in studies that recover a Rickettsiales-mitochondria affiliation⁵. An independent-endosymbioses scenario is more probable than the shared-endosymbiosis scenario: the latter requires the unlikely event of a hypothetical endosymbiotic ancestor of Rickettsiales and mitochondria escaping the host cell before engaging a new endosymbiosis that gives rise to the Rickettsiales⁵. In light of this new hypothesis, previous inferences about the nature of the mitochondrial ancestor that were based on a Rickettsiales origin, such as the requirement of pathogen-like features to survive the host cell¹⁵, the presence of flagella, *cbb₃*-type oxidases⁶, *bd*-type quinol oxidase and ATP/ADP translocase¹⁶ will have to be re-evaluated. Finally, our hypothesis stands in direct contrast to all previous hypotheses that reported an origin within the Alphaproteobacteria^{3–9,12–14}, and implies that mitochondrial endosymbiosis may be more ancient than previously thought. Our study underscores that future efforts that aim to identify the sister group of mitochondria should harness the power of cultivation-independent methods to explore alphaproteobacterial diversity in an unbiased manner.

Online content

Any Methods, including any statements of data availability and Nature Research Reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0059-5>.

Received: 29 June 2017; Accepted: 2 March 2018;

Published online 25 April 2018.

- Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
- Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
- Wang, Z. & Wu, M. An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci. Rep.* **5**, 7949 (2015).
- Williams, K. P., Sobral, B. W. & Dickerman, A. W. A robust species tree for the alphaproteobacteria. *J. Bacteriol.* **189**, 4578–4586 (2007).
- Fitzpatrick, D. A., Creevey, C. J. & McInerney, J. O. Genome phylogenies indicate a meaningful α -proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Mol. Biol. Evol.* **23**, 74–85 (2006).
- Sassera, D. et al. Phylogenomic evidence for the presence of a flagellum and *cbb₃* oxidase in the free-living mitochondrial ancestor. *Mol. Biol. Evol.* **28**, 3285–3296 (2011).
- Andersson, S. G. E. et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).
- Rodríguez-Ezpeleta, N. & Embley, T. M. The SAR11 group of alpha-proteobacteria is not related to the origin of mitochondria. *PLoS ONE* **7**, e30520 (2012).
- Ferla, M. P., Thrash, J. C., Giovannoni, S. J. & Patrick, W. M. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. *PLoS ONE* **8**, e83383 (2013).
- Darby, A. C., Cho, N.-H., Fuxelius, H.-H., Westberg, J. & Andersson, S. G. E. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet.* **23**, 511–520 (2007).
- Renvoisé, A., Merhej, V., Georgiades, K. & Raoult, D. Intracellular Rickettsiales: insights into manipulators of eukaryotic cells. *Trends Mol. Med.* **17**, 573–583 (2011).
- Thrash, J. C. et al. Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci. Rep.* **1**, 13 (2011).
- Brindefalk, B., Ettema, T. J. G., Viklund, J., Thollessen, M. & Andersson, S. G. E. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. *PLoS ONE* **6**, e24457 (2011).
- Viklund, J., Martijn, J., Ettema, T. J. G. & Andersson, S. G. E. Comparative and phylogenomic evidence that the alphaproteobacterium HIMB59 is not a member of the oceanic SAR11 clade. *PLoS ONE* **8**, e78858 (2013).
- Ball, S. G., Bhattacharya, D. & Weber, A. P. M. Pathogen to powerhouse. *Science* **351**, 659–660 (2016).
- Wang, Z. & Wu, M. Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. *PLoS ONE* **9**, e110685 (2014).
- Pittis, A. A. & Gabaldón, T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).
- Rodríguez-Ezpeleta, N. et al. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* **56**, 389–399 (2007).
- Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53**, 485–495 (2004).
- Viklund, J., Ettema, T. J. G. & Andersson, S. G. E. Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Mol. Biol. Evol.* **29**, 599–615 (2012).
- Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
- Landry, Z., Swan, B. K., Herndl, G. J., Stepanauskas, R. & Giovannoni, S. J. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *MBio* **8**, e00413–17 (2017).
- Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. metaSPAdes: a new versatile assembler. *Genome Res.* **27**, 824–834 (2017).
- Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
- Nomura, M. & Morgan, E. A. Genetics of bacterial ribosomes. *Annu. Rev. Genet.* **11**, 297–347 (1977).
- Hugenholtz, P., Skarshewski, A. & Parks, D. H. Genome-based microbial taxonomy coming of age. *Cold Spring Harbor Persp. Biol.* **8**, A018085 (2016).
- Susko, E. & Roger, A. J. On reduced amino acid alphabets for phylogenetic inference. *Mol. Biol. Evol.* **24**, 2139–2150 (2007).
- Criscuolo, A. & Gribaldo, S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
- Lang, B. F. & Burger, G. Chapter one—mitochondrial and eukaryotic origins: a critical review. *Adv. Bot. Res.* **63**, 1–20 (2012).
- Gray, M. W. Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **4**, a011403 (2012).

Acknowledgements We thank the Tara Oceans consortium for generating metagenomic datasets and for making these publicly available. We thank G. Herndl and C. Schleper for sharing metagenomic datasets before publication, and to M. Wu for sharing the genome of ‘*Candidatus Magnetococcus yuandaducum*’ before publication; K. Zaremba-Niedzwiedzka, C. Stairs, L. Erme, T. Williams, N. Lartillot, J. Alneberg, B. Quang Minh and H. C. Wang for useful advice, discussions and technical support; the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) at Uppsala University and the Swedish National Infrastructure for Computing (SNIC) at the PDC Center for High-Performance Computing for providing computational resources. This work is supported by grants of the European Research Council (ERC Starting grant 310039-PUZZLE_CELL), the Swedish Foundation for Strategic Research (SSF-FFL5) and the Swedish Research Council (VR grant 2015-04959) awarded to T.J.G.E.

Reviewer information Nature thanks T. Gabaldón, M. Gray and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.J.G.E. conceived the study. J.M., J.V. and P.O. screened metagenomic sequence datasets. J.M. and J.V. performed metagenome assemblies and metagenomic binning analyses. J.M. performed comparative genomics and phylogenetics analyses. J.M., L.G. and T.J.G.E. analysed and interpreted results. J.M. and T.J.G.E. wrote, and all authors edited and approved, the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0059-5>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to T.J.G.E.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sources of metagenomic data. Metagenomes of the four Atlantic Ocean samples were previously published²² (Sequence Read Archive (SRA) accession number SRP081826). All 243 metagenome assemblies from a previously published study²¹ were downloaded from EBI (<http://www.ebi.ac.uk/ena/about/tara-oceans-assemblies>; Supplementary Table 5). The raw sequence data from all selected Tara Oceans samples (Supplementary Table 6) were downloaded from project ERP001736 on the EBI Metagenomics portal.

RP15 pipeline. In brief, the RP15 pipeline aims to identify contigs within one or more metagenome assemblies that contain between five and fifteen ribosomal proteins that are typically encoded within the same str-spc cluster²⁵. Once a set of such 'ribocontigs' has been identified, their ribosomal proteins are aligned to the orthologous ribosomal proteins from a set of reference taxa. The separate alignments are then trimmed and concatenated into a supermatrix that is subsequently used to infer a phylogenetic tree. Because ribosomal proteins are good phylogenetic markers and bear more information than ribosomal RNA genes, the tree gives the user an estimate of the phylogenetic position of each ribocontig, and therefore of the overall phylogenetic diversity that can be found in the considered metagenome(s).

Open reading frames were predicted with Prodigal v.2.6.3³¹, ribosomal proteins were detected with PSI-BLAST v.2.3.0+³² using aligned ribosomal proteins from reference taxa as a query and aligned with MAFFT v.7.050b³³ using the L-INS-i algorithm. Alignments were trimmed with trimAl v.1.4³⁴, and phylogenetic trees were inferred with RAXML v.8.2.8³⁵ (LG + Γ model) or FastTree v.2.1.9³⁶ (LG model). The reference taxa were either a set of 90 phylogenetically diverse bacteria and archaea (the 'bacteria backbone' based on Raymann et al.³⁷), or a set of phylogenetically diverse alphaproteobacteria (74), magnetococcales (2), betaproteobacteria (4) and gammaproteobacteria (4) and mitochondria (12: *Naegleria gruberi*, *Phytophthora infestans*, *Tetrahymena thermophila*, *Monosiga brevicollis*, *Malawimonas jakobiformis*, *Hemiselmis andersenii*, *Glaucocystis nostochinearum*, *Dictyostelium discoideum*, *Cyanidioschyzon merolae*, *Chlorokybus atmophyticus*, *Bigelowia natans* and *Andalucia godoyi*) (the 'alphaproteobacteria backbone').

Metagenome assembly and binning of Atlantic Ocean samples. Each sequencing run of the four samples was processed with Trimmomatic v.0.33³⁸ to, in the stated order, trim read-through adapters (ILLUMINACLIP:TruSeq2-PE.fa:2:30:10:1:true), trim low quality base calls at the starts and ends of reads (LEADING: 3, TRAILING: 3), remove reads shorter than 60 bp (30 bp for the 5,000-m sample) (MINLEN:60) and finally remove reads that have an average phred score lower than 28 (100-m sample: 30) (AVGQUAL:28). The overall quality and presence of adaptor sequences in the unprocessed and processed reads was evaluated with FastQC v.0.11.4³⁹.

Read pairs for which both members passed the preprocessing were assembled with IDBA-UD v.1.1.1²⁴, using the *k*-mers 20, 40, 60, 80 and 100.

Upon assembly, all contigs were split in 10-kb pieces, unless the remaining fragment was smaller than 20 kb, with the CONCOCT⁴⁰ script `cut_up_fasta.py`. Then preprocessed reads of each of the four samples were mapped onto these split-contigs with the CONCOCT script `map-bowtie2-markduplicates.sh` (Bowtie2 v.2.1.0). Its output was transformed into a matrix containing differential coverage profiles for each split-contig using the CONCOCT script `gen_input_table.py`. CONCOCT v.0.40 and we then clustered the split-contigs larger than 1 kb into bins using their tetranucleotide frequency profiles and differential coverage profiles. Bins that corresponded to potential novel alphaproteobacteria were identified by checking which bins contained alphaproteobacterial ribocontigs (according to the RP15 tree). For each selected bin, we highlighted the location of their split-contigs on ESOM maps⁴¹ that were based on the tetranucleotide frequency profiles of each ≥ 5 kb split-contig of the entire assembled metagenome. The CONCOCT clustering showed a high degree of consistency with the ESOM clustering. Occasionally the split-contigs of two different CONCOCT bins would be located in the same area on the ESOM map. For such cases, the two CONCOCT bins would be merged into a single bin. Next, all bins of interest were assessed and cleaned using differential coverage, GC composition, linkage and presence of well-conserved bacterial genes⁴² information with mmgenome⁴³. Linkage information, that is, which contigs are connected through read pairs, was obtained from the read mapping with the CONCOCT script `bam_to_linkage.py` (`-fullsearch`, `--regionlength 500`). Next, split-contigs were replaced by their corresponding full-length contigs. In case not all split-contigs from a corresponding full-length contig were present in a cleaned bin, the full-length contig was only included in the bin if the majority of its split-contigs was present. Finally, the ≥ 2 kb full-length contigs of each bin were projected onto the ESOM maps. If a contig was relatively isolated from the majority of contigs on the map, that is, it exhibited a large difference in terms of tetranucleotide frequencies, it was removed from the final bin.

Metagenome assembly and binning of Pacific Ocean samples. The Tara Oceans project⁴⁴ has generated a wealth of metagenomic sequence data across 243 samples enriched for prokaryotes. Ideally, one would use all this information to acquire genomic data from novel alphaproteobacteria. This is however too computationally demanding, and thus a sub-selection of samples had to be made. To identify samples of interest, we applied the following strategy: first, we applied the RP15 pipeline (backbone: bacteria; phylogenetic inference: FastTree) on all publicly available metagenome assemblies of these samples (Supplementary Table 5). Then, for samples that the RP15 tree suggested to contain alphaproteobacterial lineages, ribocontigs phylogenetically classified as alphaproteobacteria were incorporated in another RP15 dataset (backbone: alphaproteobacteria; phylogenetic inference: RAXML) to obtain an improved phylogenetic resolution within the alphaproteobacteria. This screen showed that the large majority of the 243 samples contained novel alphaproteobacterial lineages. Finally, 45 samples (Supplementary Table 6) were chosen based on the presence of unique novel lineages and larger sequencing depth. This increases the chance to properly assemble rare, but phylogenetically interesting, novel alphaproteobacteria.

Each sequencing run of the selected samples (Supplementary Table 6) was processed with either Trimmomatic v.0.35 only or with SeqPrep (<https://github.com/jstjohn/SeqPrep>) before Trimmomatic. SeqPrep merged overlapping read pairs into longer single reads (maximum Phred score per base-call in merged read: 41) and trimmed read-through Illumina adapters (-A GATCGGAAGAGCACACG, -B AGATCGGAAGAGCGTCGT). Trimmomatic trimmed (residual) read-through Illumina adapters (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:true), trimmed low quality base-calls at the starts and ends of reads (LEADING: 3, TRAILING: 3), removed reads shorter than 60 bp (MINLEN: 60) and finally removed reads that have an average Phred score lower than 30 (AVGQUAL:30) (27, 28 or 29 for some samples; see Supplementary Table 7). Single-end mode was used for merged reads and paired-end mode for read pairs. The overall quality and presence of adaptor sequences in the unprocessed reads and different versions of processed reads were evaluated with FastQC v.0.11.4.

Although all samples have already been assembled²¹ using SOAPdenovo, we chose to re-assemble the metagenomes of '122 deep chlorophyll maximum (DCM) layer 0.22–0.45- μ m', '125 surface (SRF) layer 0.22–0.45- μ m' and '125 marine epipelagic mixed (MIX) layer 0.22–0.3- μ m' samples (Supplementary Table 6) with metaSPAdes, implemented in SPAdes v.3.7.0²³. metaSPAdes is designed specifically to deal with the metagenomic data. For the assembly the *k*-mers 21, 33, 55 and 77 were used. For 122 DCM 0.22–0.45, reads processed with SeqPrep and Trimmomatic were used as input. SeqPrep-merged reads were pooled with the Trimmomatic unpaired reads to serve as 'unpaired reads'. For 125 SRF 0.22–0.45 and 125 MIX 0.22–0.3, reads processed with Trimmomatic only were used as input. Analysis with QUAST⁴⁵ showed that, when considering only contigs of ≥ 500 bp, metaSPAdes assemblies had typically an approximately 2 \times larger total assembly size, and notably larger contigs compared with the SOAPdenovo assemblies (data not shown).

Per assembly, all contigs were split by cutting every 10 kb, unless the remaining fragment was smaller than 20 kb, with the CONCOCT script `cut_up_fasta.py`. Then kallisto v.0.42.5⁴⁶ was used to map the SeqPrep and Trimmomatic preprocessed reads from each of the 45 selected samples' sequencing runs against the split-contigs of each of the three assemblies. kallisto was run in single-end mode, treating all reads as unpaired reads. The kallisto output was transformed into a matrix containing the differential coverage profiles for each split-contig, using the `input_table.py` script that was provided by J. Alneberg (see 'Code availability'). CONCOCT v.0.4.0 then clustered the split-contigs into bins using their tetranucleotide frequency profiles and differential coverage profiles. CONCOCT was run separately for minimum split-contig length cutoffs 2 kb and 3 kb. Bins that corresponded to potential novel alphaproteobacteria were identified by checking which bins contained alphaproteobacterial ribocontigs. The bins were then assessed and cleaned using information of differential coverage, GC composition, linkage and presence of 139 genes well-conserved across Bacteria with mmgenome. Linkage information, that is, which contigs were connected through read-pairs, was obtained by mapping reads pairs processed with SeqPrep and Trimmomatic with the `map-bowtie2-markduplicates.sh` script (Bowtie2 v.2.1.0) onto the split-contigs. By default, only the bins generated with the 2-kb length cutoff were chosen, unless such a bin was deemed unsatisfactory. In that case, the corresponding bin that was generated with the 3-kb length cutoff was chosen. If multiple full-length contigs had their corresponding split-contigs distributed across two different bins, the bins were merged within mmgenome before bin cleaning. Finally, split-contigs were replaced by their full-length contigs. In case not all split-contigs from a corresponding full-length contig were present in a cleaned bin, the full-length contig was only included in the final bin if the majority of split-contigs was present.

Phylogenetic diversity in metagenome assemblies. The RP15 pipeline was used to evaluate the phylogenetic diversity present within all seven assembled samples. The pipeline was executed on all seven complete metagenome assemblies

simultaneously (backbone: bacteria, phylogenetic inference: RAXML), yielding the phylogenetic tree shown in Supplementary Fig. 1. Ribocontigs phylogenetically classified as alphaproteobacteria were incorporated in another RP15 dataset (backbone: alphaproteobacteria, phylogenetic inference: RAXML), yielding the tree shown in Fig. 1b and Supplementary Fig. 2.

Completeness and redundancy estimation. The completeness and redundancy of each final bin was estimated by checking for the presence of 139 well-conserved marker genes across bacteria. A detailed description of the method was previously published⁴⁷. In brief, it aims to provide more weight to marker genes that typically have a large distance towards other marker genes, and a smaller weight to marker genes that typically have a shorter distance towards other marker genes.

Annotation. All bins were annotated with prokka v.1.12⁴⁸, which was altered to allow for partial gene predictions on contig edges (GitHub pull request #219), with the options `--compliant`, `--partialgenes`, `--cdsrnaolap` and `--evaluate 1e-10`, and with barnap as the rRNA predictor. GC content and median intergenic space distances (only considering the coding sequence (CDS), rRNA and tRNA features as genes) were calculated with in-house scripts.

Phylogenomics dataset for alphaproteobacteria. To build a phylogenomics dataset consisting of highly conserved, vertically evolving genes that undergo a minimum of horizontal gene transfers across a phylogenetically diverse set of alphaproteobacteria, we used the following strategy: first, we selected 165 aproNOGs (available from the eggNOG database v.4.5⁴⁹) that had exactly one copy in $\geq 95\%$ of the taxa present in the aproNOGs. During the selection, the taxa *Ehrlichia ruminantium* Welgevonden, *Gluconacetobacter diazotrophicus* PA15, *Oligotropha carboxidovorans* OM5, *Bartonella bacilliformis* KC583, '*Candidatus* Hodgkinia cicadicola Dsem', '*Candidatus* Liberibacter asiaticus psy62' and '*Candidatus* Liberibacter solanacearum CLso-ZC1' were not considered, because they were exceptions that typically contained two copies or more of these genes. Because the set of taxa in the aproNOGs is limited, and lacks many phylogenetically informative taxa, and by definition Beta- and Gammaproteobacteria that are needed for the outgroup, we added orthologues from a phylogenetically diverse set of publically available alpha-, beta- and gammaproteobacteria (Supplementary Table 8), and the bins obtained in this study. Orthologue detection was done by using the trimmed alignments (available from eggNOG) of these 165 aproNOGs as queries in a PSI-BLAST search against the protein complements of to-be-added genomes. After PSI-BLAST hits were added to the original aproNOGs, single gene trees were inferred with RAXML (LG + Γ) to identify and remove all non-orthologues (paralogues, contaminations present in poorly curated genomes). Resulting clusters of orthologous genes (COGs) that did not contain ≥ 6 outgroup taxa that were monophyletic, did not have all major alphaproteobacterial taxonomic groups represented by at least one taxon, and did not contain $\geq 85\%$ of all taxa were removed. A RAXML tree was inferred from the concatenation of the remaining 90 COGs to make a final taxon selection. We aimed to reduce the number of taxa to a computationally tractable number while keeping the phylogenetic diversity as high as possible. While making single gene trees, we observed that several bins, despite coming from different samples, were virtually phylogenetically identical (tip-to-tip distance: ≤ 0.01 substitutions per site). We therefore decided to collapse these near-identical bins into 'composite bins'. A composite bin consists mostly of the phylogenetic marker genes from the bin that is most complete among the near-identical bins, but is complemented with marker genes belonging to less complete bin(s). This action simultaneously reduced the number of taxa and the percentage of missing data while keeping the phylogenetic diversity the same. We then applied a discordance filter⁵⁰ to detect COGs potentially affected by within-Alphaproteobacteria horizontal gene transfer. We removed the 20% most discordant COGs. These 72 COGs were from here on referred to as 'alphaCOGs' (Supplementary Table 2).

Phylogenomics datasets for alphaproteobacteria and mitochondria. To build a phylogenomics dataset consisting of highly conserved, vertically inherited genes that undergo a minimum of horizontal gene transfers across a diverse set of alphaproteobacteria and gene-rich mitochondrial genomes, we used the following strategy: first, we detected orthologues of all alphaCOGs in the mitochondrial genome of *A. godoyi* with a PSI-BLAST search. We chose this genome because, with 66 genes, it is currently the most gene-rich mitochondrial genome available and is present in the publicly available and curated mitoCOGs⁵¹. We then used the alphaCOG–*Andalucia* orthologue connection to merge 14 alphaCOGs with their corresponding mitoCOGs. Before merging, the mitoCOGs were filtered for the taxa *A. godoyi*, *Seculamonas ecuadoriensis*, *Histiona aroides*, *Reclinomonas americana*, *Jakoba libera*, *J. bahamiensis*, *Physcomitrella patens*, *Ostreococcus tauri*, *P. infestans* and *M. jakobiformis*. These mitochondrial genomes were chosen, because they were gene rich and exhibited short branch lengths in previous phylogenetic analyses (figure 1 and supplementary figures 5 and 6 of ref. ⁵²). By choosing genomes associated with shorter branch lengths, we aimed to minimize long branch artefacts. We then used the remaining 52 *Andalucia* mitochondrial genes to build additional orthologous groups. These genes can still be phylogenetically informative despite not being picked up by the stringent gene selection

used to build the alphaCOGs. Orthologues from all considered alphaproteobacteria and bins were detected through a BLAST search. For each of the resulting 34 orthologous groups with ≥ 70 taxa, a single gene tree was made with FastTree (LG model) to identify and remove non-orthologues (paralogues, contaminations for poorly annotated genomes). The cleaned orthologous groups were then merged with taxon-selected mitoCOGs (see above) using the mitoCOG membership information of each *Andalucia* gene. Phylogenetically near-identical bins were collapsed into composite bins as described in 'Phylogenomics dataset for alphaproteobacteria'. Next, RAXML single gene trees were inferred for all 48 orthologous groups and visually inspected. We finally selected 24 orthologous groups for which the outgroup exhibited monophyly, mitochondria exhibited monophyly, the alphaproteobacteria topology was congruent with the expected species tree (that is, were minimally affected by horizontal gene transfer) and for which the longest mitochondrial branch was $\leq 1.5\times$ longer than the longest alphaproteobacterial branch (root-to-tip lengths). These 24 orthologous groups are hereafter referred to as 'alhamitoCOGs' (Supplementary Table 2). Except for *rpl2* from *Ostreococcus*, all eukaryotic genes were mitochondrial encoded.

To evaluate the robustness of the phylogenetic placement of mitochondria recovered from the dataset described above, we built four additional phylogenomics datasets. (i) alhamitoCOGs with a more diverse mitochondrial taxon sampling. Taxa were selected as follows: for all taxa in the mitoCOGs⁵¹ that encode at least 10 out of 24 alhamitoCOGs on the mitochondrial genome, a concatenated supermatrix was constructed as described above. A phylogeny was then inferred with IQTREE (LG + F + R7, model selected by IQTREES ModelFinder⁵²). Taxa were then selected based on phylogenetic diversity, number of genes and short branch lengths (Supplementary Fig. 21). (ii) alhamitoCOGs for which the outgroups MarineProteo1, Magnetococcales, Beta- and Gammaproteobacteria were removed. (iii) alhamitoCOGs plus 14 mitochondrial-encoded genes for which mitochondrial taxa were placed on moderately longer branch lengths compared to alphaproteobacteria ($\leq 2.5\times$ longer than the longest alphaproteobacterial branch (root-to-tip lengths)). And (iv) a set of 29 nuclear encoded genes that were previously identified⁵³. To build the latter dataset, we detected their orthologues in all reference taxa (Supplementary Table 8) and bins through PSIBLAST searches, using trimmed alignments of the publically available 29 orthologous groups⁵³ (Supplementary Information, dataset 7) as query. Orthologues from *Arabidopsis thaliana*, *Cryptococcus neoformans*, *M. brevicollis*, *Nematostella vectensis*, *P. infestans* and *Spizellomyces punctatus* were directly added from this dataset. The resulting COGs were checked and cleaned for non-orthologues, and near-identical bins collapsed into composite bins as described above.

Phylogenetic inference. From the alphaCOGs and alhamitoCOGs phylogenomics datasets, three supermatrix alignments were prepared: (i) the concatenation of all aligned orthologous groups (alignment: MAFFT L-INS-i, alignment trimmer: BMGE -m BLOSUM30), (ii) alignment (i) recoded into the SR4 categories (AGNPST, CHWY, DEKQR, FILMV)²⁷ and (iii) alignment (i) trimmed with BMGE's stationary-based trimmer (BMGE -s FAST -h 0.1 -g 1). This trimmer removes sites from an alignment until all pair-wise taxa are no longer significantly compositionally heterogeneous according to Stuart's test of marginal homogeneity⁵³. Supermatrices (ii) and (iii) represent two different strategies to reduce compositional heterogeneity across taxa, a strong source of artefactual phylogenetic signal within the Alphaproteobacteria and mitochondria. An overview of all alignments (length, missing data, χ^2 score and informative sites) can be found in Supplementary Table 3. All alignments were then used for phylogenetic reconstruction under the CAT + GTR + Γ 4 model as implemented in PhyloBayes MPI v.1.7a⁵⁴. Four independent Markov chain Monte Carlo (MCMC) chains were run until convergence (maxdiff < 0.3) or a sufficient effective sample size was reached (effsize > 300). In cases in which such an effective size was not reached due to computational limitations, the MCMC distributions of the log-likelihood values, the total tree-lengths, the α parameter of the gamma distribution and the number of categories were visually inspected to assess whether a chain reached a sufficient sample size. When making consensus trees, the first 5,000 cycles (for amino acid encoded alignments) or the first 2,000 cycles (for SR4 encoded alignments) were discarded as burn-in. In addition, all non-recoded alignments were used for maximum-likelihood phylogenetic reconstruction under the LG + C60 + F + Γ 4 model (1,000 ultra-fast bootstraps) and its PMSF approximation (100 non-parametric bootstraps; guide tree: LG + F + Γ 4) as implemented by IQTREE v.1.5.0a⁵⁵. The PMSF model⁵⁶ is a computational efficient approximation of the empirical profile mixture models C10 to C60⁵⁷ that allows the user to use non-parametric bootstraps.

From each of the four additional phylogenomics datasets, we prepared a concatenated supermatrix and a stationary-trimmed concatenated supermatrix in the same manner as was done for the alphaCOG and alhamitoCOG datasets. These were then used for maximum-likelihood phylogenetic reconstruction as described above.

Posterior predictive checks. Posterior predictive checks as implemented in PhyloBayes MPI v.1.7a were performed to control whether the inferred phylogenetic

models adequately captured the across-taxa compositional heterogeneity and site-specific pattern diversity present in the alignments. In brief, it simulates alignments based on model parameter configurations sampled (here: every 50 cycles after the burn-in) from the posterior distribution (a given MCMC chain). It then calculates a set of test statistics of interest (here the maximum deviation and mean square deviation between global and taxon-specific empirical frequencies that reflects compositional heterogeneity across taxa, and the mean number of distinct characters per site that reflect site-specific pattern diversity) for each simulated alignment, yielding a null distribution per test statistic. Finally, the null distribution is compared to the observed value of the test statistic of the original alignment and a P value and a z score is calculated (one-sided test). Thus, if the observed value of the test statistic falls within the 95% confidence interval of the null distribution, the conclusion is that the property that the test statistic reflects was adequately modelled.

Progressive reduction of compositional heterogeneity. To investigate how compositional heterogeneity affects the tree topology in more detail, we executed a scheme in which we reduced the heterogeneity of a given alignment in a stepwise fashion and subsequently tracked the bootstrap supports of certain branches of interest. To do this, the unchanged supermatrices (alignments (i)) of both phylogenomics datasets were subjected to an alignment trimmer (χ^2 trimmer; see 'Code availability') that removes sites that contribute most to the overall heterogeneity (the total χ^2 score)²⁰. Here, we prepared a set of alignments with progressively more heterogeneous sites (0, 5, 10, 15, 20, 25, 30, 35 and 40%) removed, and thus with decreasing overall heterogeneity. For each resulting alignment, we inferred 100 non-parametric bootstraps under the PMSF model (guide tree: LG + F + Γ 4, mixture model: LG + C60 + F + Γ 4), implemented in IQTREE v.1.5.0a. Finally, we extracted the bootstrap supports for each bipartition of interest (see main text) and tracked them along the decreasing heterogeneity gradient. The alignment with the 20% most heterogeneous sites removed was furthermore used for a phylogenetic reconstruction with PhyloBayes-MPI under the CAT + GTR + Γ 4 model.

Topology tests. For each maximum-likelihood phylogenetic analysis that included mitochondria, we used the approximately unbiased test⁵⁸ as implemented by CONSEL v.1.20 to evaluate ten hypotheses for the phylogenetic origin of the mitochondria: sister to Rickettsiales, Alphaproteobacteria, Pelagibacterales, all non-Rickettsiales alphaproteobacteria, *Tistrella mobilis* and *Geminicoccus roseus*, *Tistrella mobilis*, *Geminicoccus roseus*, MarineAlpha9 and 10, MarineAlpha9 or MarineAlpha10. We first obtained maximum-likelihood trees with IQTREE under the constraint defined by each hypothesis (-g). We then added bootstrap trees with branch lengths (a 100 out of the 1,000 ultra-fast bootstraps in case of LG + C60 + F + Γ 4, 100 non-parametric bootstraps in case of the PMSF approximation) from the unconstrained maximum-likelihood search to improve the accuracy of the approximately unbiased test⁵⁹. Site-likelihoods for all 110 trees were then calculated under the LG + C60 + F + Γ 4 or its PMSF approximation with IQTREE (-wsl). Finally, the approximately unbiased test was performed with CONSEL v.0.20 using the IQTREE generated site-likelihoods.

Parametric simulations. A maximum likelihood tree was inferred with IQTREE from the stationary-trimmed alignment of the alphamito24 dataset under the Rickettsiales-sister constraint (-g) and the LG + C60 + F + Γ 4 model. The resulting maximum likelihood estimates of the model parameters—that is, mixture weights (+C60), observed amino acid frequencies (+F) and alpha shape parameter (+ Γ 4)—and tree were then used to simulate ten independent supermatrix alignments of the same length as the original stationary-trimmed alignment with SiteSpecific.seqgen⁵⁶, an adjusted version of SeqGen⁶⁰ able to simulate using mixture models.

Random sequence analysis. Ten independent datasets were generated by replacing all mitochondrial sequences in the alphamito24 dataset with random sequences of the same length using the EMBOSS tool makeprotseq. For each dataset, a stationary-trimmed supermatrix alignment was prepared in the same manner as for the other datasets in this study.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. input_table.py is freely available on GitHub (http://github.com/EnvGen/toolbox/tree/master/scripts/kallisto_concoct/input_table.py). alignment_pruner.pl is freely available on GitHub (https://github.com/novogit/davinciCode/tree/master/perl/alignment_pruner.pl).

Data availability. The genome bins described in this study have been deposited at DDBJ/EMBL/GenBank under the BioProject ID PRJNA390581 and whole-genome

sequencing (WGS) accessions PTJW000000000–PTLO000000000, with corresponding versions PTJW010000000–PTLO010000000. All metagenome assemblies and supermatrix alignments generated in this study are archived at the Dryad Digital Repository: <https://datadryad.org/resource/doi:10.5061/dryad.068d0d0>.

31. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
32. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
33. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
34. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
35. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
36. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
37. Raymann, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl Acad. Sci. USA* **112**, 6670–6675 (2015).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
39. Andrews, S. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> v.0.11.4 (2010).
40. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
41. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
42. Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
43. Karst, S. M., Kirkegaard, R. H. & Albersen, M. mmgenome: a toolbox for reproducible genome extraction from metagenomes. Preprint at <https://www.biorxiv.org/content/early/2016/06/15/059121> (2016).
44. Bork, P. et al. Tara Oceans studies plankton at planetary scale. *Science* **348**, 873 (2015).
45. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
46. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
47. Martijn, J. et al. Single-cell genomics of a rare environmental alphaproteobacterium provides unique insights into Rickettsiales evolution. *ISME J.* **9**, 2373–2385 (2015).
48. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
49. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
50. Williams, K. P. et al. Phylogeny of gammaproteobacteria. *J. Bacteriol.* **192**, 2305–2314 (2010).
51. Kannan, S., Rogozin, I. B. & Koonin, E. V. MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. *BMC Evol. Biol.* **14**, 237 (2014).
52. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
53. Stuart, A. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412–416 (1955).
54. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
55. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
56. Wang, H.-C., Minh, B. Q., Susko, E. & Roger, A. J. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235 (2018).
57. Si Quang, L., Gascuel, O. & Lartillot, N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323 (2008).
58. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
59. Shi, X., Gu, H., Susko, E. & Field, C. The comparison of the confidence regions in phylogeny. *Mol. Biol. Evol.* **22**, 2285–2296 (2005).
60. Rambaut, A. & Grassly, N. C. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**, 235–238 (1997).

Blastocyst-like structures generated solely from stem cells

Nicolas C. Rivron^{1,2*}, Javier Frias-Aldeguer^{1,2}, Erik J. Vrijl¹, Jean-Charles Boisset², Jeroen Korving², Judith Vivie^{2,3}, Roman K. Truckenmüller¹, Alexander van Oudenaarden², Clemens A. van Blitterswijk^{1,5} & Niels Geijsen^{2,4,5}

The blastocyst (the early mammalian embryo) forms all embryonic and extra-embryonic tissues, including the placenta. It consists of a spherical thin-walled layer, known as the trophoblast, that surrounds a fluid-filled cavity sheltering the embryonic cells¹. From mouse blastocysts, it is possible to derive both trophoblast² and embryonic stem-cell lines³, which are in vitro analogues of the trophoblast and embryonic compartments, respectively. Here we report that trophoblast and embryonic stem cells cooperate in vitro to form structures that morphologically and transcriptionally resemble embryonic day 3.5 blastocysts, termed blastoids. Like blastocysts, blastoids form from inductive signals that originate from the inner embryonic cells and drive the development of the outer trophoblast. The nature and function of these signals have been largely unexplored. Genetically and physically uncoupling the embryonic and trophoblast compartments, along with single-cell transcriptomics, reveals the extensive inventory of embryonic inductions. We specifically show that the embryonic cells maintain trophoblast proliferation and self-renewal, while fine-tuning trophoblast epithelial morphogenesis in part via a BMP4/Nodal-KLF6 axis. Although blastoids do not support the development of bona fide embryos, we demonstrate that embryonic inductions are crucial to form a trophoblast state that robustly implants and triggers decidualization in utero. Thus, at this stage, the nascent embryo fuels trophoblast development and implantation.

Although stem cells mimic development^{4–6}, a stem cell-based model of the blastocyst is lacking. To address this, we confined cultured embryonic stem (ES) cells to an average of five cells per microwell^{7,8} (Extended Data Fig. 1a). These formed non-adherent aggregates within 24 h, over which we laid trophoblast stem (TS) cells (Fig. 1a). Upon aggregation, the ES cells induced the formation of trophoblast cysts (65 h). At low frequency (less than 0.3%), cells spontaneously organized into regular TS cell cysts with internal ES cell aggregates, which we termed blastoids (Extended Data Fig. 1b).

We optimized the organization by increasing the engulfment of ES and TS cells (Extended Data Fig. 1c, d), and analysed the functions of the WNT and cAMP signalling pathways, which correlate with the formation of the blastocoel fluid-filled cavity^{9,10}. Indeed, blastocysts exposed to inhibitors of tankyrase or protein kinase A (PKA), which are used as antagonists of WNT and cAMP signalling, respectively, developed a smaller blastocoel (Extended Data Fig. 1e). Consistently, the trophoblast cells¹⁰ and TS cells produce WNT family members 6 (*Wnt6*) and 7b (*Wnt7b*), and autocrine WNT activity (Extended Data Fig. 1f, g). cAMP and WNT pathway stimulation increased TS cell cavitation and blastoid formation using several ES and TS cell lines (Fig. 1b, Extended Data Fig. 1h, i). Blastoids formed efficiently at specific ratios (70% when 8 ES cells and 20 TS cells are combined; Fig. 1c), expanded and stabilized within 65 h after seeding to diameters similar to E3.5 blastocysts (90 µm; Fig. 1d–f and Extended Data Fig. 1j).

We next analysed blastocyst transcription factors. Although the expression of pluripotency markers OCT4 (also known as POU5F1) and NANOG¹ were maintained in the pluripotent compartment (Fig. 2a, b), mRNA expression of the trophoblast marker *Cdx2*^{11,12} was low in both TS cells and blastoids compared to blastocysts (Extended Data Fig. 2a, b). The addition of the TS cell regulators FGF4 and TGFβ1¹³ increased CDX2 expression in blastoids, but levels remained low (Extended Data Fig. 2b), prompting us to seek new regulators. On the basis of observations of the STAT pathway members in blastocysts (Extended Data Fig. 2c–g), we found that IL-11 (the most abundantly expressed ligand) or 8Br-cAMP increased CDX2 protein levels^{12,13}, complemented FGF4 and TGFβ1, maintained trophoblast markers, and enhanced CDX2 expression in blastoids (Extended Data Figs. 2b, 3a–e and Fig. 2c). Blastoids remained permissive to the formation of primitive endoderm-like cells, albeit at lower numbers than blastocysts¹⁴ (Fig. 2d, Extended Data Fig. 4a–c). Cell number in the blastoid compartments was reminiscent of a mid-stage blastocyst (Extended Data Fig. 4a–c).

After injection into a foster blastocyst, ES and TS cells contribute to embryo and placental development, respectively. We analysed whether ES or TS cell lines derived de novo from blastoids had similar competences. The efficiency of TS cells derivation was similar between blastoids and blastocysts (90% generate CDX2⁺ colonies at passage 2, Extended Data Fig. 5a), whereas CDX2-low blastoids (see Methods) displayed lower derivation efficiency (35%). After injection into blastocysts, these de novo TS and ES cells contributed to the formation of the extra-embryonic tissues (embryonic day (E) 6.5 and E11.5 embryos), and epiblast, respectively (Fig. 2e, Extended Data Fig. 5a). Thus, the blastoid environment maintains the developmental potential of both lineages.

Blastoids transferred into the uterus of pseudo-pregnant mice induced deciduae formation with typical local vascular permeability (Fig. 2f), suggesting anastomosis of trophoblasts with the mother's vascular system¹⁵. Indeed, blastoid trophoblasts generated cells incorporated into the ingrowing maternal vasculature and expressing proliferin, a marker for trophoblast giant cells^{6,16} (Fig. 2g). Blastoids did not support full bona fide embryonic development, but generated numerous cells positive for a variety of extra-embryonic markers of different post-implantation trophoblast cell types^{6,15,17} (Extended Data Fig. 5b). Decidualization is a complex process regulated by the hormonal cycle and endometrium and stimulated by the embryo in rodents¹⁸. It can be partly reproduced by intraluminal deposition of sesame oil (deciduoma). We thus analysed the specificity of the decidualization process. Injection of vehicle medium alone did not induce decidualization (Fig. 2h and Extended Data Fig. 5c). Blastocysts and blastoids formed small, discrete deciduae, whereas oil triggered the formation of larger, more continuous deciduomata (Extended Data Fig. 5d). How the conceptus regulates decidualization is not well understood; however, *Aldh3a1* is one of the rare genes induced in the

¹MERLIN Institute for Technology-Inspired Regenerative Medicine, Maastricht University, Maastricht, The Netherlands. ²Hubrecht Institute for Developmental Biology and Stem Cell Research, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, Utrecht, The Netherlands. ³Single Cell Sequencing facility of the Hubrecht Institute for Developmental Biology and Stem Cell Research, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, Utrecht, The Netherlands. ⁴Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. ⁵These authors contributed equally: Clemens A. van Blitterswijk, Niels Geijsen. *e-mail: n.rivron@hubrecht.eu

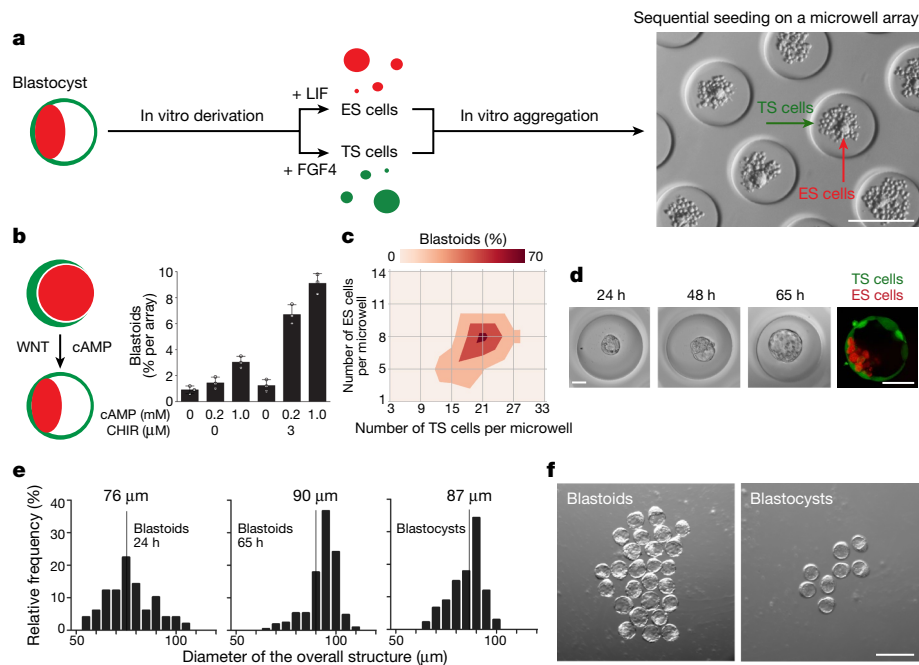


Fig. 1 | Embryonic and trophoblast stem cells form blastocyst-like structures in vitro. **a**, Schematic of blastoid formation. ES and TS cells are derived from blastocysts (left), then sequentially combined using a microwell array (right)⁷. ES cells are seeded to form non-adherent aggregates within 24 h (right, red arrow denotes a 24-h aggregate). TS cells are then added (right, green arrow denotes TS cells upon seeding). Scale bar, 200 μ m. **b**, Aggregates of ES cells engulfed by TS cells were exposed to a WNT activator (3 μ M CHIR99021; CHIR) and a cAMP analogue (0.2 or 1 mM 8Br-cAMP; cAMP). Yields of blastoids are shown (measured as the percentage of microwells containing a TS cell cyst enclosing an ES cell, 65 h after TS cell addition). 8Br-cAMP plus CHIR99021 generated significantly higher yields of 9% blastoids ($P=0.006$, two-sided Student's t -test). $n=3$ independent microwell arrays. Error bars are s.d. **c**, Blastoid

yield as a function of the initial number of TS and ES cells in individual microwells. 70% of microwells seeded with an optimal number of ES and TS cells contain a blastoid. **d**, Left, evolution of blastoid morphology from 24 to 65 h. Right, blastoids formed with ES cells positive for red fluorescent protein (RFP)-tagged histone H2B (H2B-RFP⁺ ES cells; red) and TS cells positive for green fluorescent protein (GFP) (GFP⁺ TS cells; green). Scale bars, 50 μ m. **e**, Distribution of the overall diameters of blastoids, at 24 and 65 h, and of blastocysts collected on day E3.5. Bar plots indicate the percentage frequency of specific diameters. $n=50$ independent blastoids or blastocysts. Vertical line denotes the median. **f**, Light microscopy image showing blastoids and E3.5 blastocysts (high magnification images shown in Extended Data Fig. 1j). Scale bar, 200 μ m. All experiments were repeated at least three times with similar results.

decidua by the implanting conceptus¹⁸. Consistently, ALDH3A1 was locally expressed in a group of decidual cells in the mesometrial side of blastocyst- and blastoid-induced deciduae (Fig. 2i and Extended Data Fig. 5e). Thus, the blastoid recapitulates key aspects of uterine

implantation (discrete decidualization, vascular anastomosis, and patterned ALDH3A1 expression in the decidua).

Comparing transcriptomes using unsupervised clustering analysis (Fig. 3a) confirmed that blastoids resembled blastocysts at E3.5, when

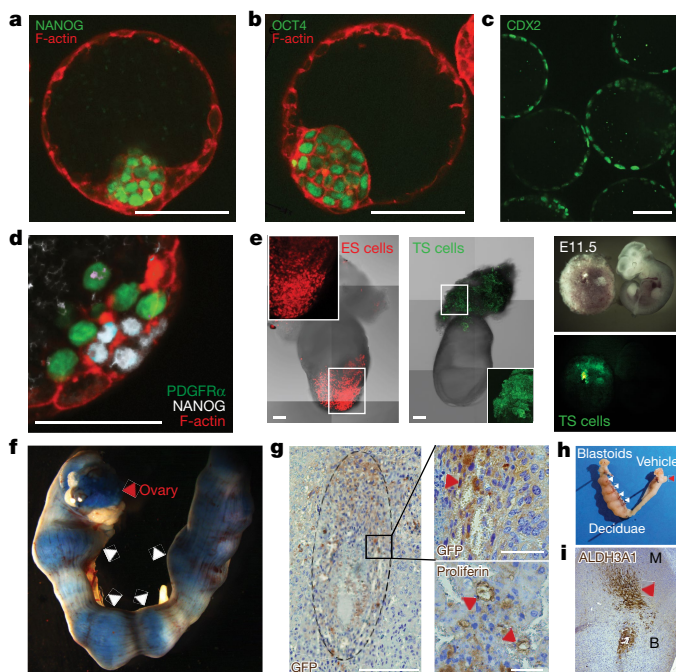


Fig. 2 | Blastoids implant in utero and trigger the formation of patterned deciduae. **a**, **b**, Immunofluorescent staining for NANOG (green, **a**) and OCT4 (green, **b**) in blastoids counterstained with F-actin (red). **c**, Immunofluorescent staining for CDX2 within blastoids (Extended Data Fig. 2b). **d**, Immunofluorescent staining for NANOG (white) in blastoids formed with ES cells comprising a primitive endoderm-specific PDGFR α -H2B-GFP reporter (green) and counterstained with F-actin (red). Scale bars, 50 μ m (**a–d**). **e**, E6.5 and E11.5 embryos from blastocysts injected with de novo blastoid-derived ES cells (left) and TS cells (middle and right). Insets show the contribution of H2B-RFP⁺ ES cells to the epiblast (left) and of GFP⁺ TS cells to the extra-embryonic tissue (middle). Also see Extended Data Fig. 5a. **f**, Uterus transferred with blastoids at E3.3–E3.5 and explanted at E7.5. Mice are injected systemically with Evan blue dye, revealing typical local vascular permeability of the implantation sites (deciduae: white arrowheads; ovary: red arrowhead). **g**, Left, anti-GFP staining of a decidua containing an in utero developed blastoid formed with GFP⁺ TS cells (E6.5). Scale bar, 1 mm. Right, anti-GFP staining (top; scale bar, 100 μ m) and proliferin (bottom; scale bar, 50 μ m) within histological sections of deciduae including vascular lumens (red arrowheads). **h**, Uterus transferred with blastoids (left horn) and vehicle medium (right) at E3.3–E3.5, and explanted at E6.5 (deciduae: white arrowheads; ovary: red arrowhead). **i**, Anti-ALDH3A1 staining of a representative decidua induced by a blastoid. B, blastoid implantation site; M, mesometrial side. Red arrowhead denotes the decidua sub-population expressing ALDH3A1. All experiments were repeated at least three times with similar results.

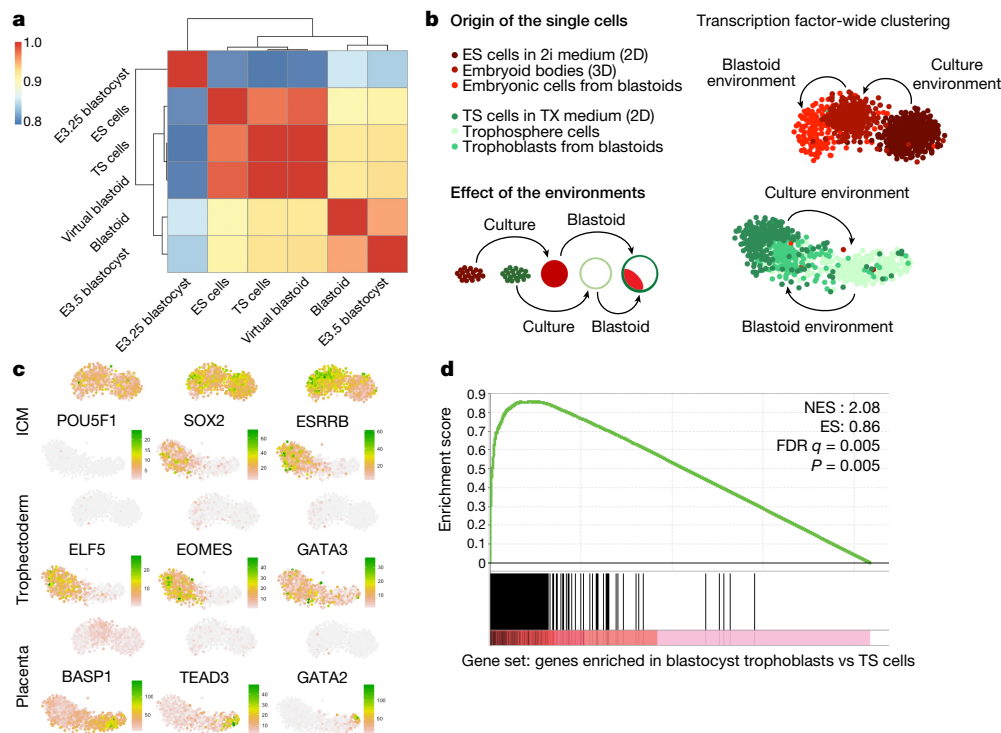


Fig. 3 | Communication between blastoid embryonic and trophoblast compartments shifts the transcriptome towards an E3.5 blastocyst state. **a**, Bulk sequencing. Whole-transcriptome distance map. The total number of either blastocysts or blastoids is 50. **b–d**, Single-cell RNA sequencing. **b**, *t*-distributed stochastic neighbour embedding (*t*-SNE) map representation of transcriptome similarities between 1,577 cells collected from 2D-cultured parental ES cells in serum-free (2i) medium (316 cells) and parental TS cells in serum-free (TX) medium (290 cells), blastoid cells (367 cells), TS and ES cells cultured separately in blastoid medium within microwells (that is, 336 trophosphere cells, and 268 embryoid body cells). Arrows indicate transcriptome shifts due either to the culture environment (for example, culture medium and microwell confinement) or to the blastoid environment (for example, ES and TS cell communication and confinement within the blastoid cyst). Clustering is

based on the transcription factors expressed in the blastocyst. See whole-transcriptome clustering in Extended Data Fig. 6c, d. The embryonic or trophoblast blastoid cells are identified based on FACS-sorting indexes. **c**, *t*-SNE map representation of key transcription factors for the ICM, trophoctoderm and placenta. **d**, Comparison of trophoblasts from blastoids and blastocysts. Gene set enrichment analysis. Genes are ranked according to their difference in expression between trophoblasts of blastoids (left of the horizontal axis) and parental TS cells (right of the horizontal axis). Black bars depict the position of a gene set of 281 genes significantly enriched in trophoblasts of blastocyst compared to parental TS cells ($P = 0.05$, see Supplementary Table 1 sheet 5). This gene set is significantly enriched in the trophoblasts of blastoids (normalized enrichment score (NES) = 2.08, $P = 0.005$). ES, enrichment score; FDR, false discovery rate.

lineage transcriptomes are stabilized¹⁹, consistent with the committed state of TS and ES cells. Similar clustering occurred when limited to transcription factors, and dissimilarities were primarily related to metabolism (Extended Data Fig. 6a, b and Supplementary Table 1 sheet 1). Importantly, the transcriptome of blastoids distinctly separated from their parental ES and TS cell lines (Fig. 3a). We then combined in silico the transcriptome of TS and ES cells cultured separately in blastoid medium: this virtual blastoid clustered with the parental cell lines, not with the blastoids (Fig. 3a), suggesting that the communication between the embryonic and trophoblast cells contributed to the E3.5-like transcriptome shift. Comparison with the blastoid transcriptome identified differences within cell cycle, epithelial junctions and MAPK signalling pathways (Extended Data Fig. 6b, Supplementary Table 1 sheet 2), which all drive blastocyst formation^{2,20–22}. A role for the TGF β signalling pathway also surfaced, the contribution of which to blastocyst formation is less clear^{23,24}.

We further evaluated the relative contribution of the culture condition (for example, culture medium, microwell confinement) and blastoid environment (for example, ES and TS cell communication, blastoid cyst confinement) to the transcriptome shift via single-cell sequencing of 2D-cultured parental TS and ES cells, blastoid cells, and TS and ES cells cultured individually in blastoid culture environment. In such conditions, TS cells alone formed trophospheres and ES cells formed embryoid bodies. Non-supervised clustering analysis clearly assessed the separation of embryonic cells (ES cells, blastoid embryonic cells and embryoid body cells) and trophoblast cells

(TS cells, blastoid trophoblast cells and trophosphere cells) (Fig. 3b and Extended Data Fig. 6c–e). Within the embryonic or trophoblast cluster, cells changed mainly owing to the culture environment (for example, culture medium and microwell confinement) and to the blastoid environment (for example, communication between embryonic and trophoblast cells, and cyst confinement). All embryonic cells maintained core pluripotency markers and did not express epiblast stem-cell markers (Fig. 3c and Extended Data Fig. 6g). Consistent with the de novo derivation of stem-cell lines, the core transcription factors of the TS cells were also maintained in the blastoid environment (Fig. 3c). In sharp contrast, the absence of embryonic cells strongly decreased these transcription factors in the cells of trophospheres, which expressed differentiation genes typical of post-implantation placental cell types (Fig. 3c and Extended Data Fig. 6f, g).

Consistent with the morphogenesis of a cyst, blastoid trophoblasts enriched in epithelial transcripts (Extended Data Fig. 6f, g and Supplementary Table 1 sheets 4–6), the proteins of which were correctly localized (Extended Data Fig. 7a–d). Finally, gene set enrichment analysis revealed that blastoid trophoblasts were largely enriched in the transcripts of blastocyst trophoblasts (Fig. 3d, Supplementary Table 1 sheet 5). Thus, the blastoid environment prevents the differentiation of trophoblasts, while fuelling the epithelial morphogenesis of a trophoctoderm-like cyst.

To analyse functionally the embryonic inductions regulating trophoctoderm development, we tested the specificity of the compartments' interactions by substituting the ES cells for other cell types,

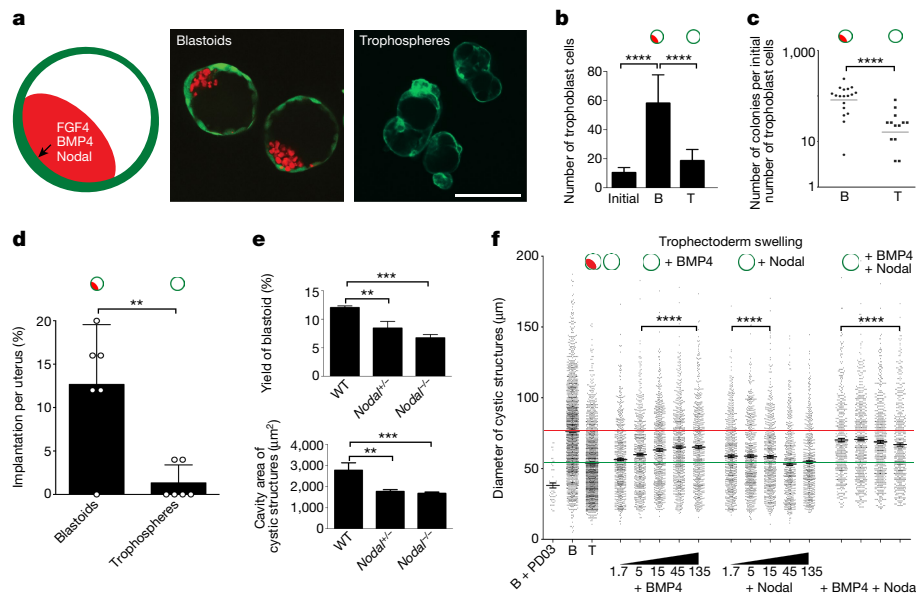


Fig. 4 | Embryonic inductions regulate trophoblast proliferation, self-renewal, epithelial morphogenesis and implantation, partially via BMP4 and Nodal. **a**, Blastoids formed with GFP⁺ TS cells and H2B-RFP⁺ ES cells are compared with trophospheres formed with GFP⁺ TS cells alone. Scale bar, 100 μ m. Images are representative of three experiments. **b**, **c**, Quantification of embryonic inductions previously assessed in blastocysts. The schemes above the graphics depict blastoids or trophospheres. **b**, Trophoblast proliferation. Count for the initial number of TS cells seeded (see Extended Data Fig. 1a) and present after 65 h within blastoids (B) and trophospheres (T). Horizontal bars indicate the mean number of trophoblasts. **** $P = 2.10^{-5}$, two-sided Student's *t*-test. $n = 30$ blastoids or trophospheres. Error bars indicate s.d. See Extended Data Fig. 10b. **c**, Trophoblast self-renewal. Colony formation unit (CFU) determined as the number of CDX2⁺ colonies divided by the mean number of trophoblasts in blastoids or trophospheres. Horizontal bars denote mean CFU. **** $P = 4.10^{-5}$, two-sided Student's *t*-test.

$n = 13$ independent blastoids or trophospheres. Error bars indicate s.d. **d**, Trophoblast implantation. Percentage of implantation sites in uteri transferred with blastoids or trophospheres at E3.3–E3.5, and explanted at E7.5. $n = 6$ independent mice. Horizontal bars indicate mean percentage. ** $P = 0.003$, Student's two-sided *t*-test. Error bars indicate s.d. **e**, **f**, Morphogenetic functions of TGF β activators. **e**, Yield (top) and cavity area (bottom) of blastoids formed with wild-type (WT), *Nodal*^{+/-} and *Nodal*^{-/-} ES cells. Horizontal bars denote mean yield or cavity area. **** $P = 2.10^{-4}$, two-sided Student's *t*-test. $n = 3$ independent microwell arrays. Error bars indicate s.d. **f**, Overall diameter of blastoids, trophospheres and trophospheres exposed to BMP4, Nodal or their combination. Horizontal bars indicate mean of $n = 250$ independent structures. Error bars denote s.e.m. Combinations of BMP4 and Nodal are (in order): 5 ng ml⁻¹ BMP4 + 5 ng ml⁻¹ Nodal; 45 ng ml⁻¹ BMP4 + 45 ng ml⁻¹ Nodal; 45 ng ml⁻¹ BMP4 + 5 ng ml⁻¹ Nodal; 45 ng ml⁻¹ BMP4 + 45 ng ml⁻¹ Nodal.

including mouse epiblast stem cells (EpiSCs, in vitro analogues of E5.75 embryos). EpiSCs induced the formation of fewer blastoids (fivefold) of smaller size (Extended Data Fig. 8a). Other cell types did not support trophoblast development, demonstrating that specific inductions emerge from ES cells. FGF4, originating from the inner cell mass (ICM), regulates trophoblast proliferation and self-renewal^{2,25,26}. Consistently, ES cells expressed eightfold more *Fgf4* mRNA than EpiSCs (Extended Data Fig. 8b), and induced transcriptional signatures for MAPK signalling activity and an enhanced cell cycle in blastoid trophoblasts (Extended Data Fig. 8c, d and Supplementary Table 1 sheet 7). Accordingly, there were twice as many trophoblasts in blastoids than trophospheres, and blastoid trophoblasts formed more (fivefold) CDX2⁺ colonies than trophosphere trophoblasts (Fig. 4b, c). We then tested whether morphogenesis of the epithelial cyst, an important transformation within blastoids, is functionally regulated by resident ES cells. Consistent with induction of trophoblast morphogenesis genes, ES cell titration within blastoids increased both blastoid cavitation and diameter (Extended Data Fig. 8e, f). Finally, as the trophoblast state regulates the decidualization response^{16,18}, we transferred blastoids and trophospheres in utero. Consistent with their post-implantation transcriptome signature, trophospheres had diminished potential for decidualization as compared to blastoids (sixfold, Fig. 4d). Thus, resident embryonic cells functionally maintain trophoblast proliferation and self-renewal, and prevent trophoblast differentiation into post-implantation placental cell types, while shaping their epithelial architecture and implantation potential.

The TGF β signalling pathway is active in the blastocyst^{23,24}. However, its functions are unknown owing to delayed detection of the defects

after loss-of-function mutations (E5.5–E6.5)²³. This delay may reflect a developmental robustness rooted in functional redundancy of ligands (for example, Nodal–activin–BMP), plasticity of signalling pathways, and/or technical limitations of detection²³. Also, the temporal overlap of lineage commitment and morphogenesis in early blastocysts limits the interpretation of compartment-specific inducible genetic models. The blastoid system overcomes these limitations by permitting (i) genetic and physical uncoupling of the trophoblast and embryonic compartments, and (ii) disentangling blastocyst morphogenetic processes from earlier lineage commitment events.

In the blastoid, the embryonic cells induced TGF β signalling pathway activity in trophoblasts (Extended Data Fig. 8c, d and Supplementary Table 1 sheets 6, 7). Also, the RNAs of the TGF β activators *Bmp4* and *Nodal* are largely restricted to the embryonic cells (Extended Data Fig. 9a). We concluded that BMP4 and Nodal produced by the embryonic compartment induce TGF β activity in the trophoblast.

To assess loss-of-function in the context of other inductions (for example, FGF4), we generated ES cells with a heterozygote or homozygote *Nodal* knockout (Extended Data Fig. 8b). These cells had a decreased capacity to form blastoids, which had smaller cavities (Fig. 4e). Next, we explored TGF β gain-of-function activators. Stimulation of trophospheres with BMP4 and Nodal regulated a similar number of genes (30% overlap). Combined, they upregulated expression of the transcription factors *Cdx2*, *Id2* and *Klf6*, induced WNT activity, which functionally regulates cavitation (see Supplementary Table 1 sheet 8 and Fig. 1b), and downregulated the expression of *Cldn4* and *Krt8*, which are overexpressed in differentiated trophospheres (Extended Data Fig. 9c–f, Supplementary Table 2 sheets 1–3). As for morphogenetic features, Nodal increased the cavitation

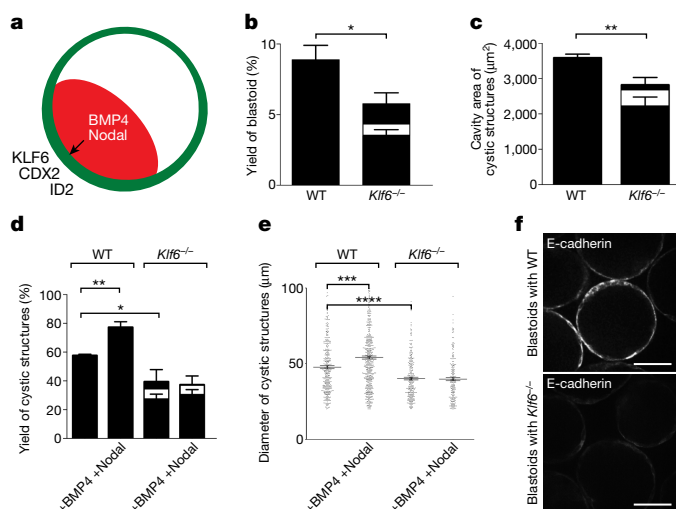


Fig. 5 | KLF6 mediates the response to BMP4 and Nodal, and trophoblast morphogenesis. **a**, Blastoid formation and stimulation of trophospheres with BMP4 and Nodal upregulates expression of KLF6, CDX2 and ID2 in trophoblasts. **b**, Yield of blastoids formed with wild-type and *Klf6*^{-/-} TS cells (two clones). Horizontal bars indicate mean yield. **P* = 0.01, two-sided Student's *t*-test. *n* = 3 independent microwell arrays. **c**, Cavity area of blastoids formed with wild-type and *Klf6*^{-/-} TS cells (two clones). Horizontal bars denote the mean cavity area. ***P* = 0.005, two-sided Student's *t*-test. *n* = 3 independent microwell arrays. **d**, Yield of trophospheres formed with wild-type and *Klf6*^{-/-} TS cells, and after stimulation with 45 ng ml⁻¹ BMP4 and 5 ng ml⁻¹ Nodal. Horizontal bars indicate mean yield. **P* = 0.02, ***P* = 0.001, two-sided Student's *t*-test. *n* = 3 independent microwell arrays. Error bars in **b–d** denote s.d. **e**, Diameters of trophospheres formed with wild-type and *Klf6*^{-/-} TS cells, and after stimulation with 45 ng ml⁻¹ BMP4 and 5 ng ml⁻¹ Nodal. Horizontal bars indicate mean diameter. ****P* = 0.001, *****P* = 0.0001, two-sided Student's *t*-test. *n* = 80. Error bars are s.e.m. **f**, Immunostaining for E-cadherin (DECMA antibody) in blastoids formed with wild-type and *Klf6*^{-/-} TS cells. Images are representative of three independent experiments. Scale bars, 50 μm.

of trophospheres and had a milder effect on their diameter, whereas BMP4 induced the opposite effect. BMP4 and Nodal together increased both the cavitation (Extended Data Figs. 9g, 10a) and diameter (120%, Fig. 4f) of trophospheres. As there was no change in cell numbers (Extended Data Fig. 10b), we concluded that BMP4 and Nodal regulate cavitation and swelling, consistent with the regulation of epithelial components. Accordingly, TGFβ signalling pathway inhibition using LDN193189 reduced the size of the blastocoel cavity in blastocysts (Extended Data Fig. 10c). Thus, BMP4 and Nodal contribute to trophoblast epithelial morphogenesis.

Blastoid formation and the stimulation of trophospheres with BMP4 and Nodal both upregulated *Klf6* in trophoblasts (Fig. 5a, Supplementary Table 2 sheets 1–3). Consistent with a continuum reflecting trophoblast and placenta development, trophospheres strongly expressed *Klf6*, which is somehow involved in early placenta development²⁷ (Extended Data Fig. 6g). We thus generated two *Klf6*^{-/-} TS cell lines (Extended Data Fig. 10d) and analysed trophoblast morphogenesis. In 2D culture, the lines appeared similar to the parental line (morphology, proliferation, and levels of E-cadherin; Extended Data Fig. 10e). However, mirroring the ES cell *Nodal* knockout phenotype, both TS cell lines had a reduced capacity to form blastoids, which had a smaller cavities (Fig. 5b, c). This defect was partly TS cell autonomous, as their capacity to form trophospheres was also reduced (Fig. 5d, e). In addition, contrary to their parental line, *Klf6*^{-/-} trophospheres did not respond to TGFβ activators (Fig. 5d, e, Extended Data Fig. 10f). *Klf6* has been linked to epithelial functions and E-cadherin expression²⁸. Indeed, E-cadherin and *Krt8* were downregulated within *Klf6*^{-/-} blastoids (Fig. 5f, Extended Data Fig. 10g). Thus, *Klf6* is an important target of BMP4 and Nodal, regulating the epithelial morphogenesis of the trophoblast.

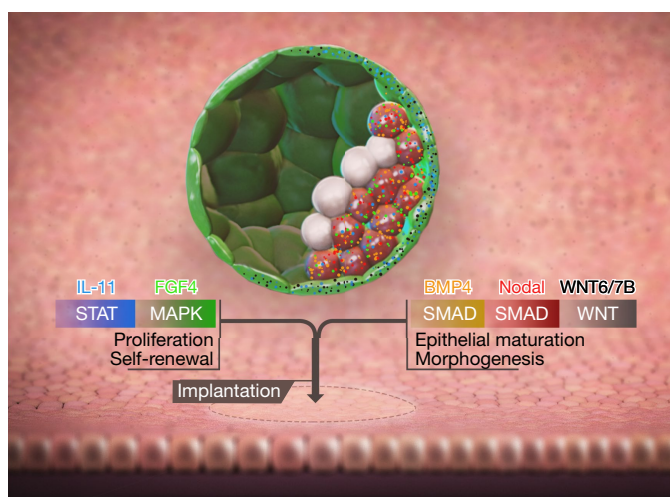


Fig. 6 | Embryonic inductions drive trophoblast development and implantation. Left, trophoblast proliferation and self-renewal. FGF4 largely originates from the embryonic compartment. IL-11 is the most abundantly expressed STAT regulator, and originates from both the embryonic and trophoblast compartments. Together, FGF4 and IL-11 regulate trophoblast proliferation and self-renewal, and CDX2 expression. Right, trophoblast epithelial maturation and morphogenesis. BMP4 and Nodal largely originate from the embryonic compartment and, besides maintaining the trophoblast stem-cell state¹³, regulate trophoblast epithelial maturation and morphogenesis. Among WNT ligands¹⁰, WNT6 and WNT7B are expressed in the trophoblast, and regulate epithelial morphogenesis. Altogether, signals originating from the embryonic cells fuel trophoblast proliferation, morphogenesis and generate a trophoblast state prone to in utero implantation.

Assigning functional roles to compartment interactions in blastocysts is challenging owing to (i) the difficulty to form blastocysts without ICM cells (trophospheres)^{26,29}, (ii) the temporal overlap between lineage commitment and blastocyst morphogenesis, and (iii) the relative speed of blastocyst development, which limits the interpretation of compartment-specific inducible models. Here, we describe the formation of blastoids, which morphologically and transcriptionally resemble E3.5 blastocysts, recapitulate key features of trophoblast development, and implant in utero. This model overcomes many limitations of blastocyst research and proposes new mechanisms of embryonic inductions that drive trophoblast development and in utero implantation (Fig. 6).

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0051-0>.

Received: 21 September 2015; Accepted: 28 February 2018;
Published online 2 May 2018.

- Wennekamp, S., Mesecke, S., Nédélec, F. & Hiiragi, T. A self-organization framework for symmetry breaking in the mammalian embryo. *Nat. Rev. Mol. Cell Biol.* **14**, 452–459 (2013).
- Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).
- Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **433**, 519–523 (2008).
- van den Brink, S. C. et al. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 (2014).
- Harrison, S. E., Sozen, B., Christodoulou, N., Kyprianou, C. & Zernicka-Goetz, M. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science* **356**, eaal1810 (2017).
- Rai, A. & Cross, J. C. Three-dimensional cultures of trophoblast stem cells autonomously develop vascular-like spaces lined by trophoblast giant cells. *Dev. Biol.* **398**, 110–119 (2015).

7. Rivron, N. C. et al. Tissue deformation spatially modulates VEGF signaling and angiogenesis. *Proc. Natl Acad. Sci. USA* **109**, 6886–6891 (2012).
8. Vrij, E. et al. Directed assembly and development of material-free tissues with complex architectures. *Adv. Mater.* **28**, 4032–4039 (2016).
9. Manejwala, F., Kaji, E. & Schultz, R. M. Development of activatable adenylate cyclase in the preimplantation mouse embryo and a role for cyclic AMP in blastocoel formation. *Cell* **46**, 95–103 (1986).
10. Kemp, C., Willems, E., Abdo, S., Lambiv, L. & Leyns, L. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Dev. Dyn.* **233**, 1064–1075 (2005).
11. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev. Biol.* **313**, 614–629 (2008).
12. McDole, K. & Zheng, Y. Generation and live imaging of an endogenous Cdx2 reporter mouse line. *Genesis* **50**, 775–782 (2012).
13. Kubaczka, C. et al. Derivation and maintenance of murine trophoblast stem cells under defined conditions. *Stem Cell Reports* **2**, 232–242 (2014).
14. Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. & Hadjantonakis, A.-K. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081–3091 (2008).
15. Simmons, D. G., Fortier, A. L. & Cross, J. C. Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. *Dev. Biol.* **304**, 567–578 (2007).
16. Red-Horse, K. et al. Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. *J. Clin. Invest.* **114**, 744–754 (2004).
17. Latos, P. A. & Hemberger, M. From the stem of the placental tree: trophoblast stem cells and their progeny. *Development* **143**, 3650–3660 (2016).
18. McConaha, M. E., Eckstrum, K., An, J., Steinle, J. J. & Bany, B. M. Microarray assessment of the influence of the conceptus on gene expression in the mouse uterus during decidualization. *Reproduction* **141**, 511–527 (2011).
19. Ohnishi, Y. et al. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16**, 27–37 (2014).
20. Gotoh, N. et al. The docking protein FRS2alpha is an essential component of multiple fibroblast growth factor responses during early mouse development. *Mol. Cell. Biol.* **25**, 4105–4116 (2005).
21. Saba-El-Leil, M. K. et al. An essential function of the mitogen-activated protein kinase Erk2 in mouse trophoblast development. *EMBO Rep.* **4**, 964–968 (2003).
22. Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. & Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc. Natl Acad. Sci. USA* **95**, 5082–5087 (1998).
23. Papanayotou, C. & Collignon, J. Activin/Nodal signalling before implantation: setting the stage for embryo patterning. *Phil. Trans. R. Soc. Lond. B* **369**, 1–8 (2014).
24. Mesnard, D. & Constam, D. B. Imaging proprotein convertase activities and their regulation in the implanting mouse blastocyst. *J. Cell Biol.* **191**, 129–139 (2010).
25. Gardner, R. L. Flow of cells from polar to mural trophectoderm is polarized in the mouse blastocyst. *Hum. Reprod.* **15**, 694–701 (2000).
26. Gardner, R. L., Papaioannou, V. E. & Barton, S. C. Origin of the ectoplacental cone and secondary giant cells in mouse blastocysts reconstituted from isolated trophoblast and inner cell mass. *J. Embryol. Exp. Morphol.* **30**, 561–572 (1973).
27. Matsumoto, N. et al. Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6. *Blood* **107**, 1357–1365 (2006).
28. DiFeo, A. et al. E-cadherin is a novel transcriptional target of the KLF6 tumor suppressor. *Oncogene* **25**, 6026–6031 (2006).
29. Tarkowski, A. K. & Wróblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol.* **18**, 155–180 (1967).

Acknowledgements We would like to thank J. Deschamps for providing the *Cdx2*-eGFP mice; M. Creighton for providing V6.5 mouse ES cells; V. Prideaux, J. Garner and J. Rossant for providing the F₄ mouse TS cell lines; A.-K. Hadjantonakis for providing the PDGFR α -H2B-GFP mouse ES cells; N. Schelling for helping with logistics; S. van der Elst for helping with FACS assays; A. de Graaf for helping with microscopes; L. Yvernogeu for helping with the imaging of embryos; B. El Haddouti for helping with uterus injections; H. Begthel for helping with histology; W. de Lau for helping with the TCF luciferase assay; C. Rabouille and H. Clevers for commenting on the manuscript. N.C.R. is grateful for support from The Netherlands Organisation for Scientific Research (NWO-ZonMw) Translational Adult Stem Cell grant (1.6005008). We apologize that all relevant publications could not be cited.

Author contributions N.C.R. conceived and directed the project, conducted the experiments and wrote the manuscript. J.F.-A. performed the single-molecule FISH assay, the qRT-PCR assay, contributed to generating the CRISPR knockout lines and to the RNA-sequencing assays. E.J.V. performed the high-content imaging assays. J.K. completed the blastocyst complementation assays and the in utero transfers. J.V. helped prepare the libraries of single cells for RNA sequencing. J.-C.B. contributed to the RNA-sequencing assays. R.K.T. contributed to the design of the microwell array. A.v.O. contributed to the RNA-sequencing assays. C.A.v.B. and N.G. helped to direct the project and contributed equally to the study. All the authors discussed and corrected the manuscript.

Competing interests N.C.R., E.J.V., C.A.v.B. and N.G. are inventors on the patent US14/784,659 and PCT/NL2014/050239 (April 2014).

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0051-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0051-0>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to N.C.R.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Culture of ES cells. Experiments were done using the following cell lines: V6.5, H2B-RFP V6.5 sub-clone, PDGFR α -H2B-GFP, Sox17-GFP and IB10. The V6.5 cell line was derived from C57BL/6 \times 129/Sv background and obtained from the laboratory of R. Jaenisch. The IB10 cell line was a subclone of the E14 cell line (129/Ola) and obtained from the laboratory of H. Clevers. The PDGFR α -H2B-GFP cell line was derived from an ICR background and obtained from the laboratory of A.-K. Hadjantonakis. A Sox17-GFP cell line was derived from unknown background and obtained from the laboratory of S. Morrison. All cells were cultured in 2i conditions, meaning on gelatin-coated plates in B27N2 medium complemented with leukaemia inhibitory factor (LIF, 10 ng ml⁻¹), PD0325901 (1 μ M, AxonMed 1408) and CHIR99021 (3 μ M, AxonMed 1386) as developed previously³. Cells were routinely passaged every two days for 15 days before being used for blastoid formation.

Culture of TS cells. Experiments were done using the F₄, F₁ and Cdx2-eGFP cell lines. The TS cell lines F₄ (ICR \times ICR) were obtained from the laboratory of J. Rossant. The F₁ (CBA \times C57BL/6) and Cdx2-eGFP TS cell lines were derived by N. Rivron according to the methods described by previously². Cdx2-eGFP TS cell lines were derived from a mouse reporter line in which eGFP is fused to the endogenous locus of *Cdx2*, as described previously¹². Cells were then cultured on Matrigel in phenol red-free TX medium, a serum-free medium developed previously¹³. TX medium contains DMEM/F12 (phenol red-free, with L-glutamine), L-ascorbic-acid-2-phosphate (64 μ g ml⁻¹), sodium selenite (14 ng ml⁻¹), insulin (19.4 μ g ml⁻¹), sodium bicarbonate (543 μ g ml⁻¹), holo-transferrin (10.7 μ g ml⁻¹), penicillin streptomycin, FGF4 (25 ng ml⁻¹), TGF β 1 (2 ng ml⁻¹) and heparin (1 μ g ml⁻¹). Alternatively, and when mentioned, TS cells were cultured in serum-rich conditions (TS medium) as described previously². Cells were routinely passaged every 4 days before being used for blastoid formation.

Culture of other cell types. Mouse EpiSCs were in cultured serum-free conditions, as previously described³⁰. Human ES cells were cultured in E8 medium. C2C12 and COS7 cells were culture in DMEM (Gibco, phenol red-free with L-glutamine), supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich).

Microwell arrays. Microwell arrays were formed as previously described^{7,8} and inserted into 12-well plates or directly imprinted into 96-well plates. Each microwell array included 1,000 or 400 microwells, respectively, with a diameter of 200 μ m, as described in Extended Data Fig. 1a.

Culture of blastoids. The full protocol was repeated independently in the Hubrecht and MERLN institutes, and is available at Protocol Exchange³¹ and on <http://www.blastoid.org>. ES cells were seeded by dispensing a cell suspension in DMEM medium (Gibco 31966021) with non-essential amino acids (Gibco, 11140-050), β -mercaptoethanol (Gibco, 21985-023), 10% FBS (Sigma-Aldrich) and LIF (10³ U ml⁻¹), on top of the microwell arrays with a cell concentration resulting in the pooling of a mean of five cells per microwell. Within 24–36 h, the ES cells formed tight, round aggregates. TS cells were then seeded on top of the ES cells aggregates at a concentration resulting in the pooling of a mean of 12 cells per microwell. Upon settling of the cells within the microwells, TX medium was added and complemented with Y27632 (20 μ M, AxonMed 1683), CHIR99021 (3 μ M, AxonMed 1386), 8Br-cAMP (1 mM, Biolog Life Science Institute B007E), FGF4 (25 ng ml⁻¹, R&D systems 5846F4), TGF β 1 (15 ng ml⁻¹, Peprotech 100-21), IL-11 (30 ng ml⁻¹, Peprotech 200-11) and heparin (1 μ g ml⁻¹). The time of TS cell seeding is considered as the starting point (0 h). Within 24 h, the TS cells aggregated with the ES cells. At 24 h, 1 mM 8Br-cAMP was added to the medium. Within 48 h, 2–8% of the aggregates formed a cavity, which expanded and stabilized by 65 h. A blastoid is defined based on the morphological parameters of E3.5 blastocysts, as a cystic structure with an outer circularity superior to 0.9 (circularity = $4\pi(\text{area}/\text{perimeter}^2)$), and a diameter comprised between 70 and 110 μ m, including a single regular cavity lined by a single layer of TS cells and including ES cells. Cystic structures refers to all the TS-cell cavitated structures with a diameter of cavity greater than 20 μ m. WNT3A-conditioned medium was obtained from the laboratory of H. Clevers; WNT3A recombinant protein was obtained from Cell Guidance Systems. XAV939, a tankyrase inhibitor (Tocris Bioscience), is used as an antagonist of WNT signalling and acts via stimulation of β -catenin degradation and stabilization of axin; XAV939 was used at 15 μ M.

Trophospheres were obtained by seeding TS cells (mean of 12 cells per microwell) in the same medium as used for blastoids. TX medium without growth factors (blank medium) refers to medium containing DMEM/F12 (phenol red-free, with L-glutamine), L-ascorbic-acid-2-phosphate (64 μ g ml⁻¹), sodium selenite (14 ng ml⁻¹), insulin (19.4 μ g ml⁻¹), sodium bicarbonate (543 μ g ml⁻¹), holo-transferrin (10.7 μ g ml⁻¹) and penicillin streptomycin.

TCF luciferase reporter assay. A TOP/TK-Renilla reporter plasmid system was used for the detection of β -catenin driven *Wnt*-transcriptional activity, as previously described³². The TOP reporter construct contains three optimal copies of T cell factor (TCF)/lymphoid enhancer factor (LEF) transcription factor sites

upstream of a thymidine kinase minimal promoter that, when bound by β -catenin induces transcription of the luciferase reporter gene. The *Renilla* reporter construct contains thymidine kinase-Renilla luciferase (TK-Renilla), which drives strong WNT-independent activity of the *Renilla* gene and serves as a measure of cell viability. Luciferase reporter gene analysis was performed in TS cells cultured in 2D: 16,000 cells were seeded into each well of a 96-well plate in TX culture condition. After 24 h, the cells were transiently co-transfected using Lipofectamine 3000 (Invitrogen, L3000001). Cells were stimulated 16 h later with TX or blastoid culture medium and the Porcupin inhibitor IWP2 (2.5 μ M). TX without growth factors was used as a negative control and WNT3A-conditioned medium was used as a positive control. After 24 h, cells were lysed using passive lysis buffer and luciferase and *Renilla* activity was measured with the Dual Luciferase Reporter Assay System (Promega, E1910). Each condition was performed in quadruplicate and the reporter activity was expressed as mean \pm s.d.

Screening for regulators of Cdx2 in TS cells. The TS cell line used for this assay is homozygous for Cdx2-eGFP and seeded at 20,000 cells per cm² on Matrigel-coated plates, in TX medium including 25 ng ml⁻¹ FGF4 and 2 ng ml⁻¹ TGF β 1. After 24 h, the medium was replaced by TX medium without FGF4 and TGF β 1 (blank medium) and candidate molecules were added. PD0325901 (1 μ M) was used as a negative control to downregulate CDX2 expression (see Extended Data Fig. 2). Cells were assessed for eGFP expression 48 h after the addition of the candidate molecules, using a BD FACSCalibur. Data were analysed using FlowJo. The basal level of Cdx2-eGFP expression was defined by gating the population of TS cells cultured for 48 h in blank medium (see Extended Data Fig. 3).

Immunohistochemistry. Immunofluorescence was performed as described by the laboratory of J. Rossant (<http://lab.research.sickkids.ca/rossant/lab-resources/>). Antibodies were used against CDX2 (MU392A-UC; 1:400 dilution), Nanog (Abcam ab84447; dilution 1:200), OCT4 (Santa Cruz 5279; 1:200 dilution), GATA6 (R&D Systems AF1700; 1:200), PDGFR α (R&D Systems AF1062; 1:150 dilution), ELF5 (Santa Cruz sc-9645; 1:250 dilution), MASH2 (also known as ASCL2; Genetex GTX60272; 1:250 dilution), TEAD4 (Abcam ab58310; 1:400 dilution) and proliferin (E10, Santa Cruz sc-271891; 1:250 dilution), TPBPA (Abcam ab104401; 1:250 dilution), placenta lactogen (P17, Santa Cruz, sc-34713; 1:200 dilution), HAND1 (Abcam ab115256; 1:250 dilution), and E-cadherin clone DECMA-1 (Sigma U3254; 1:500 dilution). All images were taken using a PerkinElmer Ultraview VoX spinning disk microscope combined with a Leica SP8.

Derivation of de novo ES and TS cell lines from blastoids and injection into blastocysts. Blastoids cultured for 65 h and freshly isolated E3.5 blastocysts (E3.5 CBA \times C57BL/6) were individually transferred into wells of 96-well plates containing mouse embryonic fibroblasts. TS cell lines were derived in TS medium containing 20% FBS, 25 ng ml⁻¹ FGF4, 2 ng ml⁻¹ TGF β 1 and 1 μ g ml⁻¹ heparin, as described previously². ES cell lines were derived in B27N2 medium supplemented with 10 ng ml⁻¹ LIF, 1 μ M PD0325901 and 3 μ M CHIR99021 (2i medium). The derivation was considered successful if colonies appeared on passage two after blastocyst or blastoid plating. ES or TS cells (12–15 cells) were injected into E3.5 blastocysts from C57BL/6 mice. The injected blastocysts were transferred into the uterus of CBA \times C57BL/6 pseudo-pregnant females. Embryos were explanted from the uterus and dissected from the deciduae at E6.5.

Uterus transfer. Pseudo-pregnant F₁ females were selected for oestrous and placed with vasectomized males in the evening. Mice were considered to mate at midnight, and the next morning, plugged females were separated. Blastoids and trophospheres were cultured and selected as described above, and 25 blastoids or trophospheres were injected into both horns of the uterus of day E3.3–E3.5 pseudo-pregnant Bl6/CBA females. The blastoids and trophospheres were transferred on the ovary side of the uterus, using a medium containing LIF. Uteri were explanted on day E6.5 or E7.5, fixed in 4% paraformaldehyde and processed for paraffin embedding and histology.

qRT-PCR. RNA was obtained using TRIzol reagent (Thermo Scientific 15596-018) extraction method. Retrotranscription was performed using the Superscript III kit (Invitrogen 18080-044). Quantitative PCR was performed using IQ SYBR Green Supermix (Bio-Rad 1708880) and 30 ng of RNA using the following gene-specific primers: *Cdx2*: 5'-AAAGTGAGCTGGCTGCCACACTTG-3', 5'-TCCATCAGTAGATGCTGTTCTGCTGG-3'; *Tcfap2c*: 5'-GAGGTGCAGAAATGTGGACGA-3', 5'-CCCCAAAGGGTCTCTGGTCA-3'; *Tead4*: 5'-TTGAGCGAAGCTTCCAGGAG-3', 5'-TTCCGACCATACATCTTGCT-3'; *Gata3*: 5'-GCTCCTTGC TACTCAGGTGAT-3', 5'-GGAGGGAGAGAGGAATCCGA-3'; *Eomes*: 5'-TGATCATCACCACAGAGGGC-3', 5'-ACTGTGTCTCTGAGAAGGTG-3'; *Elf5*: 5'-TTGCTCGCAGAGGTTACTCC-3', 5'-TGTTCCGGCTGTGACAGTCTT-3'; *Hand1*: 5'-GCCTACTTGATGGACGTGCT-3', 5'-TGCTGAGGCACTCCCTTTT-3'; *Cdh1*: 5'-CCAAGACAGTATCAGGGTCA-3', 5'-ACTGCTGGTCAGGATCGTTG-3'; *Id2*: 5'-CCTGCATCACCAGAGACCTG-3', 5'-GGGAATTCAGATGCTGCAA-3'; *Sox2*: 5'-GATCAGCATGTACCTCCCCG-3', 5'-CTGGGCCATGTGCAGTCTAC-3'; *Actb*: 5'-TGTCGAGTCGCGTCCACC-3', 5'-TCGTCATCCATGGCGAACTGG-3'.

RNA sequencing. Blastocysts were flushed from the uterus of CBA \times C57BL/6 E3.5 mice and sorted as E3.25 or E3.5 blastocysts according to their phenotype: early blastocysts had a smaller number of cells, a smaller diameter, a smaller blastocoel and a more prominent ICM than late blastocysts. Blastoids were formed using F₄ wild-type or GFP⁺ TS cells and V6.5 wild-type or H2B-RFP ES cells. For the bulk sequencing, blastocysts or blastoids ($n = 50$) were pooled into TRIzol. The transcriptional profile of virtual blastoids was determined by combining in silico, at a 1:3.5 ratio, ES and TS cells cultured in 2D, in blastoid medium, for 65 h. For single-cell sequencing, blastoids or Tyrod's acid treated, zona-free, blastocysts were sequentially exposed to 0.3 mg ml⁻¹ collagenase IV (Gibco, 17104019) and a 1:3 dilution of Tryple Express Select X10 (ThermoFisher A1217701), and subsequently dissected with glass capillaries of different diameters ranging from 100 to 20 μ m. Single cells were FACS-sorted into 384-well plates as described previously³³. For this analysis, we analysed a total of 1,577 cells including 290 parental TS cells cultured in 2D in TX medium; 316 parental ES cells cultured in 2D in 2i medium; 367 blastoid cells; 336 TS cells cultured alone in blastoid medium within microwells, which formed trophospheres; 268 ES cells cultured alone in blastoid medium within microwells, which formed embryoid bodies; and 60 blastocyst cells extracted from three different E3.5 blastocysts.

CEL-Seq library preparation. For the bulk sequencing experiment (Fig. 3a), the whole RNA from blastocysts and blastoids was extracted from TRIzol. For cell lines, 20 ng of total extracted RNA was used as starting material, as measured by Qubit RNA assay (Life Technologies). All samples were processed using CEL-Seq2 protocol³⁴, the CEL-Seq1 primers³⁵ and the Life Technologies Ambion kit (AM1751) as previously described³⁶. For the single-cell sequencing experiments, cells were processed according to the SORT-seq method, as previously described³³ and clustered using the Race-ID method, as previously described³⁶. *t*-SNE map representation of transcriptome similarities was based on 1,577 cells collected from parental ES cells cultured in 2D in 2i medium (316 cells), parental TS cells cultured in 2D in TX medium (290 cells), blastoid cells (367 cells), TS cells cultured in blastoid medium within microwells (that is trophospheres, 336 cells), ES cells cultured in blastoid medium within microwells (that is embryoid bodies, 268 cells). Libraries were sequenced on an Illumina Nextseq 500 using 75-bp high output paired-end sequencing. Transcription factors were selected according to the Riken transcription factor database, and lowly expressed ones were filtered out before plotting the heat maps. Differential expression analysis was done using the DESeq package³⁷, and pathway enrichment analysis was done using DAVID³⁸ and Gorilla³⁹.

Gene set enrichment analysis. Analysis was performed according to the standard procedure of gene set enrichment analysis (GSEA) analysis (<http://software.broadinstitute.org/gsea/>). The gene set was established by extracting genes that were differentially regulated between the cells of the trophoderm of blastocysts (27 cells) and TS cells (33 cells). The pre-ranked list of genes differentially expressed between the trophoblasts of blastoids (59 cells) and TS cells (33 cells) was established by extracting all the genes that were differentially regulated between clusters established using the RaceID clustering. These populations were defined based on the RaceID clustering. The gene set consisted of 281 genes differentially regulated and with $P < 0.05$.

Curated gene sets. Curated gene sets were used to identify, within the list of GSEA enriched genes, genes related to the TGF β signalling pathway or to epithelial cells (highlighted in red and green, respectively, in Supplementary Table 1 sheet 5). The gene set for the TGF β pathway was composed of the GSEA gene sets: reactome signalling by TGF β receptor complex; KEGG TGF β signalling pathway; Biocarta TGF β pathway, and PID SMAD2/3 nuclear pathways. The gene set for the epithelial cells was composed of the GO gene sets: GO:0002066; GO:0002065; GO:0090136; GO:0002064; GO:0030855; GO:0072148; GO:0002070; GO:0003382; GO:0050673; and GO:0060429.

Whole-mount single-molecule FISH. The Affymetrix ViewRNA ISH Cell Assay Kit was used to perform whole-mount single-molecule FISH on freshly isolated E3.5 blastocysts. The probe used was designed to hybridize against *Id2* (catalogue VB6-10967), *Bmp4* (catalogue VB1-13681), *Nodal* (catalogue VB6-18786), and *Il11* (catalogue VB6-19190-VC) mRNA. Blastocysts were imaged in steps of 0.5- μ m

using a 63 \times objective on a PerkinElmer Ultraview VoX spinning disk microscope combined with a Leica SP8.

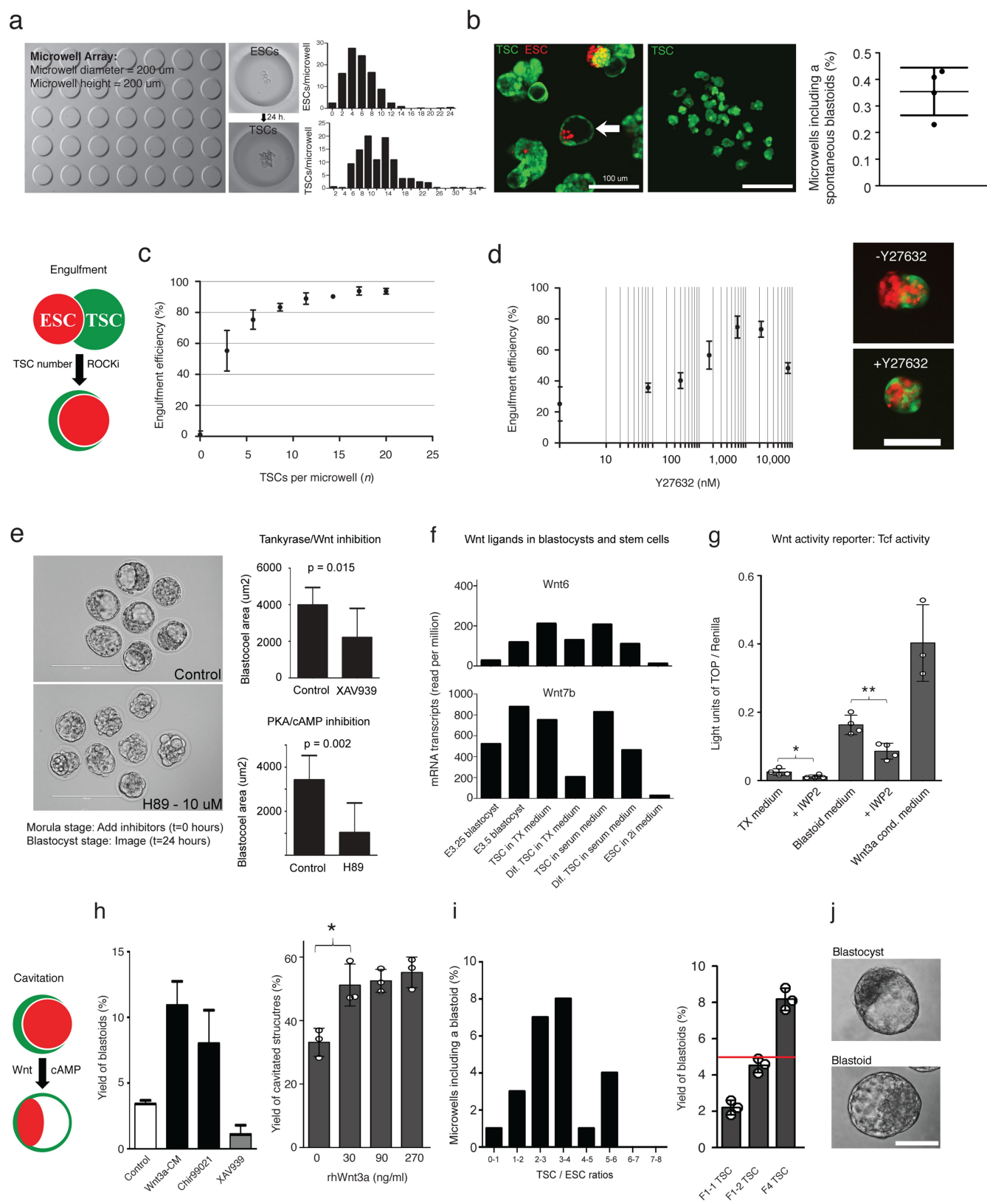
BMP4 and Nodal screens. The screen was performed by adding proteins within the blastoid medium at both 0 and 48 h. Images were acquired over 800 microwells per condition. The yield of blastoids (Fig. 4b) is measured as described in the 'Culture of blastoids' section. The yield of cystic structures (Figs. 4d and 5) is measured as the percentage of microwells that contained a structure with a cavity that has a maximal diameter $>20 \mu$ m. The diameter of cystic structures (Fig. 4e) is measured by image analyses using the microscope measurement tool, Fiji, by measuring the larger diameter of the larger cavity within each microwell. The circularity is measured using Fiji, as follows: (circularity = $4\pi(\text{area/perimeter}^2)$). Combinations of BMP4 and Nodal were: 5 ng ml⁻¹ BMP4 + 5 ng ml⁻¹ Nodal; 5 ng ml⁻¹ BMP4 + 45 ng ml⁻¹ Nodal; 45 ng ml⁻¹ BMP4 + 5 ng ml⁻¹ Nodal; 45 ng ml⁻¹ BMP4 + 45 ng ml⁻¹ Nodal.

CRISPR knockout. Two guide RNAs (gRNAs) were designed to flank the first exon of *Nodal* or the first two exons of the *Klf6* gene. ES and TS cells were transfected with a plasmid for expression of Cas9 protein, gRNAs and the puromycin-resistance gene. ES cells were transfected in 2D feeder-free conditions; TS cells were transfected in suspension for 4 h. Both ES and TS cells were subsequently plated on mouse embryonic fibroblasts. Puromycin was added 24 h after transfection to the culture media to apply selection during the following 48 h. Colonies were allowed to grow until they were able to be picked, then screened for deletion of the region of interest.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Data generated during the study are available in the Supplementary Tables and in the Gene Expression Omnibus (GEO) public repository under accession GSE99786.

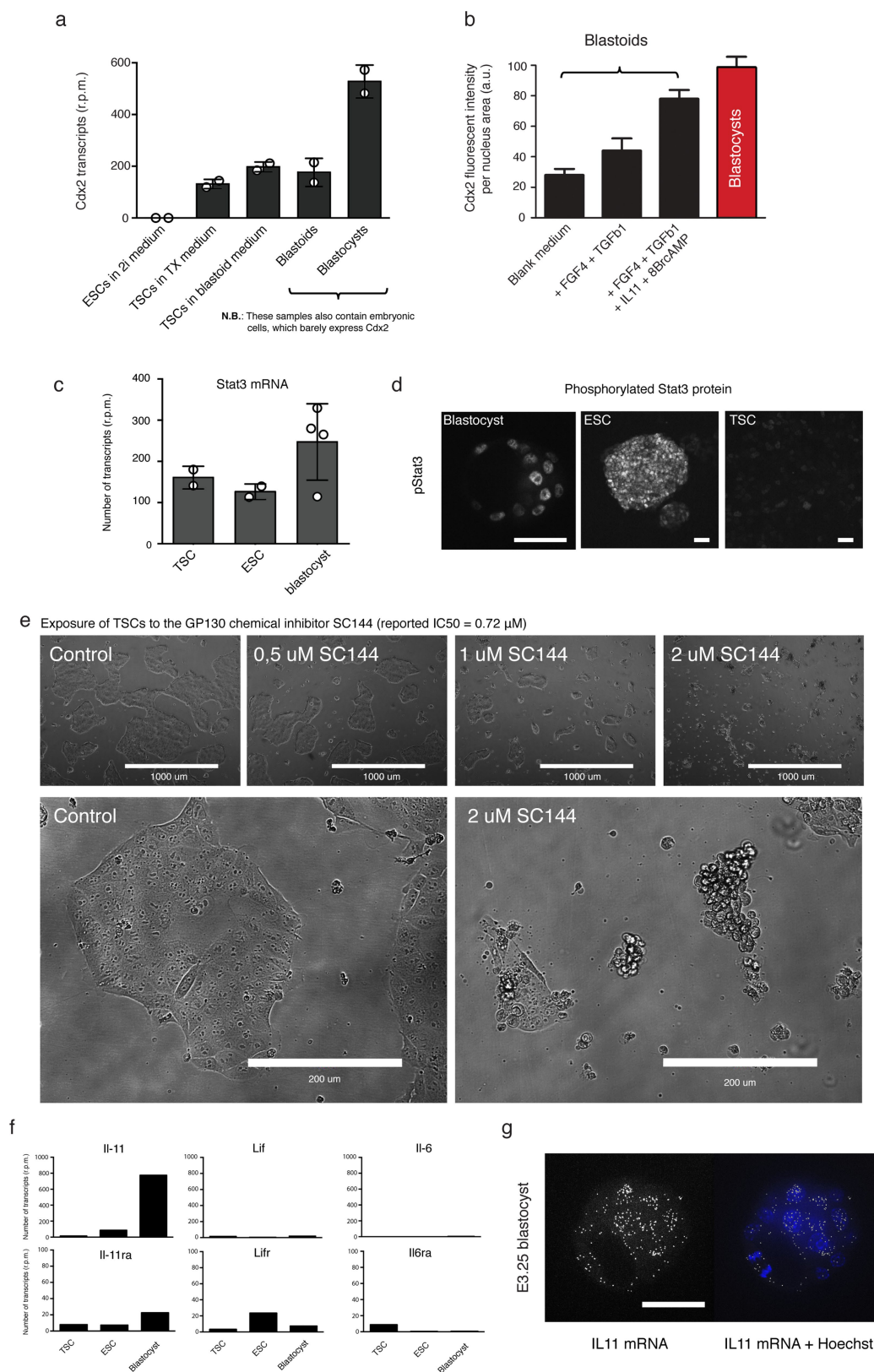
30. Tesar, P. J. et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
31. Rivron, N. C. In vitro generation of blastoids from trophoblast stem cells and embryonic stem cells. *Protoc. Exch.* <https://doi.org/10.1038/protex.2018.051> (2018).
32. van de Wetering, M. et al. The β -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).
33. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e3. (2016).
34. Hashimshony, T. et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
35. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Reports* **2**, 666–673 (2012).
36. Grün, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
37. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
38. Dennis, G., Jr et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
39. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. Gorilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
40. Vrij, E. J. et al. 3D high throughput screening and profiling of embryoid bodies in thermoformed microwell plates. *Lab Chip* **16**, 734–742 (2016).
41. Nakamura, T. et al. SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Res.* **43**, e60 (2015).
42. Kolodziejczyk, A. A. et al. Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485 (2015).
43. Qiu, D. et al. Klf2 and Tfcp2l1, two Wnt/ β -catenin targets, act synergistically to induce and maintain naive pluripotency. *Stem Cell Reports* **5**, 314–322 (2015).
44. Morgani, S. M. et al. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3**, 1945–1957 (2013).
45. Hussein, S. M., Duff, E. K. & Sirard, C. Smad4 and β -catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of *Msx2*. *J. Biol. Chem.* **278**, 48805–48814 (2003).
46. Labbé, E. et al. Transcriptional cooperation between the transforming growth factor- β and Wnt pathways in mammary and intestinal tumorigenesis. *Cancer Res.* **67**, 75–84 (2007).



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | In vitro formation of blastoids. a, Non-adherent hydrogel microwell arrays were formed by replica molding using polydimethylsiloxane (PDMS) stamps as previously described^{7,8,40}. The array used for 12-well plates contains 1,000 cylindrical structures of 200 μm diameter and height. Upon cell seeding, owing to the non-adherent properties of the hydrogel, all cells slide into the microwells. Upon settling, the number of ES cells per microwell follows a distribution. Mean number per microwell: 5.2; half of the microwells contained between 4 and 6 cells per microwell (right, top). Upon aggregation of ES cells (24–36 h depending on the cell line), TS cells were seeded. Mean number per microwell: 12; half of the microwells contained between 10 and 14 cells (right, bottom; time point 0 h). **b,** After culture in blank medium (see Methods), TS and ES cells mostly arranged into non-organized structures forming trophoblasts cysts. TS cells cultured without ES cells (65 h) formed fewer cystic structures than a co-culture. Rarely, TS cells enclosed ES cells and formed regular cystic structures morphologically similar to the blastocyst (white arrow). $n = 4$ independent microwell arrays. Structures shown in **b** were taken out of the microwell array at 65 h. Scale bars, 100 μm . **c, d,** Optimization of the engulfment of ES cells by TS cells. **c,** Dosing the number of TS cells. ES cells were seeded at $t = -24/36$ h. At $t = 0$ h, different numbers of TS cells were seeded on top of the ES cell aggregates. After 24 h, the engulfment efficiency was defined by measuring the percentage of coverage of TS cells around ES cells. The most efficient yields were observed when more than 12 TS cells were added to ES cell cultures. $n = 250$ independent microwells. The centre value is the mean, error bars are s.d. **d,** Left, optimization of the concentration of Y27632. At $t = 0$ h, different concentrations of Y27632 were added. At $t = 24$ h, the engulfment efficiency was measured as the percentage of coverage of TS cells around ES cells. The most efficient yields were obtained with 20 μM Y27632. $n = 250$ independent microwells. The centre value is the mean. Right, optimized engulfment. Images of ES cells (red) engulfed by TS cells (green) using the optimized conditions

(mean of 12 TS cells per microwell and 20 μM Y27632, 80% engulfment efficiency). Images are representative of three independent experiments. Scale bar, 100 μm . Error bars are s.d. **e,** Blastocoel area of blastocysts formed from E2.5 CBA \times C57BL/6 morula, selected for initiation of compaction and cultured in M16 medium for 24 h along with antagonists of WNT (XAV939, 15 μM), PKA (H89, 10 μM) or DMSO (1:1,000, control). $n = 10$ independent blastocysts. $P = 0.015$ and $P = 0.002$, two-sided Student's t -test. The centre values are medians. Errors bars are s.d. **f,** RNA-sequencing data for *Wnt6* and *Wnt7b*, in E3.25 and E3.5 blastocysts, TS cells (TSC) and ES cells (ESC). TS cells were cultured in serum-free (TX) or serum-rich (TS) medium (see Methods). Differentiation was induced by the removal of growth factors. **g,** TCF luciferase assay for WNT activity in TS cells (see Methods). WNT secretion was blocked using IWP2 (2.5 μM). $n = 4$ independent biological samples of TS cell culture. $*P = 0.045$, $**P = 0.0057$, two-sided Student's t -test. The centre values are medians. Errors bars are s.d. **h,** Induction of cavitation. Blastoids were defined as described in the Methods. ES and TS cells were seeded in serum-free TX medium including Y27632 (20 μM) and a WNT modulator. WNT3A-conditioned medium (50% of the total volume), CHIR99021 (3 μM), or the combination of CHIR99021 (3 μM) and XAV939 (15 μM) was added at the time of TS cells seeding ($t = 0$ h). $n = 3$ independent microwell arrays. $*P = 0.017$, two-sided Student's t -test. The centre values are medians. Errors bars are s.d. **i,** Yield of blastoids depending on the initial ratio of TS to ES cells, at $t = 0$ h, within individual microwells (left). Yield of blastoids formed using three lines of TS cells representative of the scope of efficiency observed (right). Different lines were isolated upon CBA \times C57BL/6 matings (F_{1-1} , F_{1-2} , derived by N.R.) and ICR \times ICR matings (F_4 , provided by J. Rossant). $n = 3$ independent microwell arrays. The centre values are the mean. Error bars represent s.d. The red line represents the median of the three cell lines (5% of the total number of microwell per array). **j,** Bright-field images of a representative E3.5 blastocyst (top) and a blastoid (bottom). Scale bar, 50 μm .



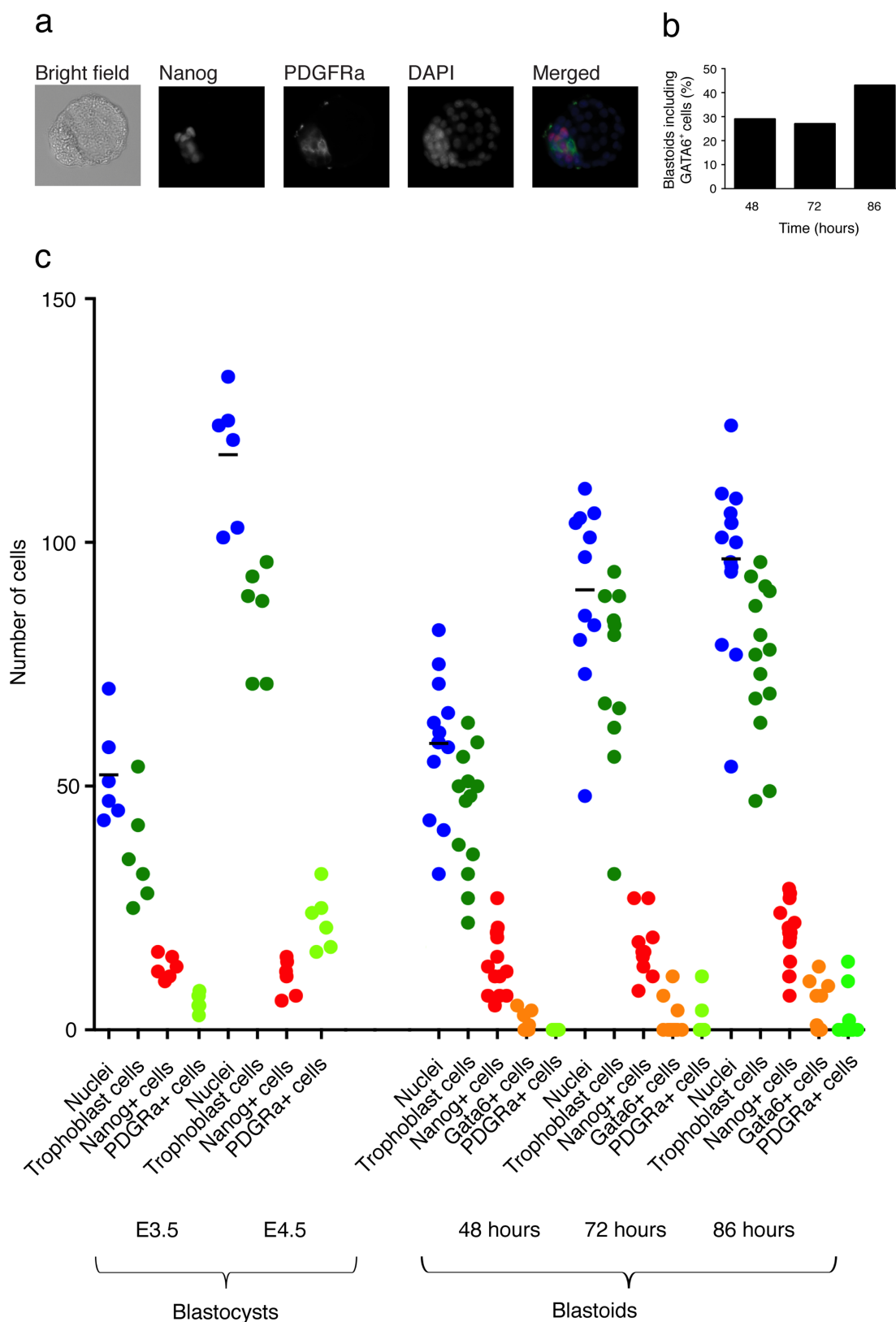
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Regulation of trophoblast cells. **a**, The number of *Cdx2* transcript reads per million mapped reads (r.p.m.) as measured in bulk samples of: ES cultured in 2i medium and TS cells cultured in TX medium, TS cells cultured in blastoid medium, and blastoids and blastocysts. Note that the blastoids and blastocysts comprise both trophectoderm and ICM cells, the latter expressing only very low levels of *Cdx2*. $n = 2$ independent biological samples (see Methods). **b**, Measurement of CDX2 fluorescent intensity in blastoids and blastocysts. Blastoids were fixed and stained with an anti-CDX2 antibody (see Methods). E3.5 blastocysts were used as a positive control. $n = 15$ independent blastoids or blastocysts. Error bars are s.d. **c**, The number of

Stat3 transcripts per million reads as measured in bulk samples of ES and TS cells and blastocysts. **d**, Immunostaining for phosphorylated STAT3 in a representative blastocyst colony of ES and TS cells (see Methods). Scale bar, 50 μm . **e**, TS cells cultured in the presence of different concentrations of the STAT/GP130 pathway inhibitor SC144. Both growth and viability were affected by concentrations of at least 1 μM . **f**, The number of transcripts per million reads for the STAT pathway ligands *Il11*, *Lif* and *Il6*, along with their receptors, as measured in bulk samples of ES cells, TS cells and blastocysts. **g**, Whole-mount single-molecule FISH for *Il11*, in an E3.25 blastocyst. Scale bar, 50 μm . See Methods for further details.

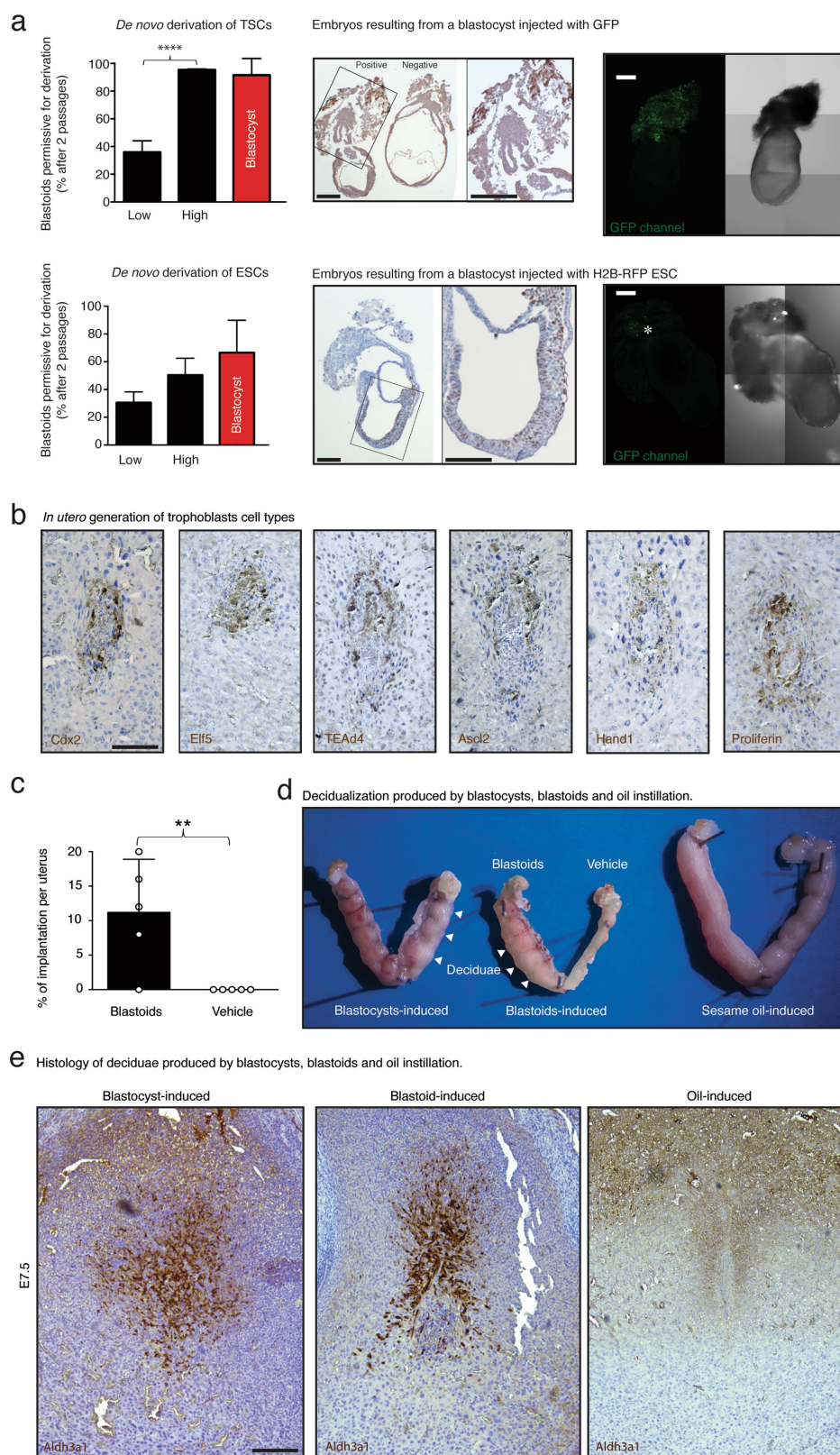
Extended Data Fig. 3 | Screen for regulators of CDX2 expression in TS cells. **a**, Calibration of the cell lines. The assay is performed using CDX2-eGFP⁺ TS cells and as described in the schematic and the Methods. For the initial calibration of the assay, gating was set so that wild-type TS cells do not appear in the gate (non-specific fluorescence, left FACS plot). In that condition, after 48 h, the following appeared in the gate: 50% of the CDX2-eGFP⁺ cells cultured in blank TX medium; 88% of the CDX2-eGFP⁺ TS cells cultured in TX medium including FGF4 (25 ng ml⁻¹) and TGFβ1 (2 ng ml⁻¹); 10% of the CDX2-eGFP⁺ TS cells cultured in TX medium including FGF4 (25 ng ml⁻¹), TGFβ1 (2 ng ml⁻¹) and PD0325901. Biological triplicates show very minor variability (*n* = 3 independent biological samples, right). Error bars are s.d. Scale bar, 100 μm. **b**, Calibration of the assay and primary screen. For the primary screen of proteins and small molecules, the condition with TX medium including FGF4 and TGFβ1 was used as a positive control. The gating was set up such that 50% of these cells appear in the gate. The condition with

TX medium without FGF4 and TGFβ1 was used as a negative control, and 20% of these cells appear in the gate. 8Br-cAMP (0.04 to 5 mM), IL-11 (4 to 500 ng ml⁻¹), LIF (3 to 375 ng ml⁻¹), BMP4 (1 to 125 ng ml⁻¹), and IGF2 (1 to 125 ng ml⁻¹) were added to the medium. The value is the measurement of a single sample. The typical s.d. for this assay is shown in **a**. **c**, Secondary screen: combinations of hits. 8Br-cAMP (1 mM), IL-11 (30 ng ml⁻¹), or 8Br-cAMP (1 mM) + IL-11 (30 ng ml⁻¹) were added to TX medium including FGF4 and TGFβ1. The full blastoid medium was also tested (see Methods). **d**, **e**, Markers of multipotency, differentiation and epithelization in stimulated TS cells. TS cells were grown for 48 h in blank medium (see Methods), TX medium (including FGF4 and TGFβ1) or TX medium supplemented with 8Br-cAMP and IL-11. Representative bright-field images of TS cells grown for 48 h in TX medium or in blastoid medium are shown (**d**), along with gene expression as characterized using bulk RNA sequencing (**e**, *n* = 3). Scale bars, 1,000 μm.



Extended Data Fig. 4 | Evolution of the number of cells in the blastoid compartments. **a**, E3.5 and E4.5 blastocysts ($n = 6$) were immunostained using Nanog and PDGFR α antibodies, and nuclei were counterstained with DAPI (see Methods). **b**, Percentage of blastoids including GATA6⁺ cells from a total number of 14 blastoids. **c**, Number of cells in each of the blastoid or blastocyst compartments, namely trophoblast, embryonic and primitive endoderm, were counted based on an immunostaining

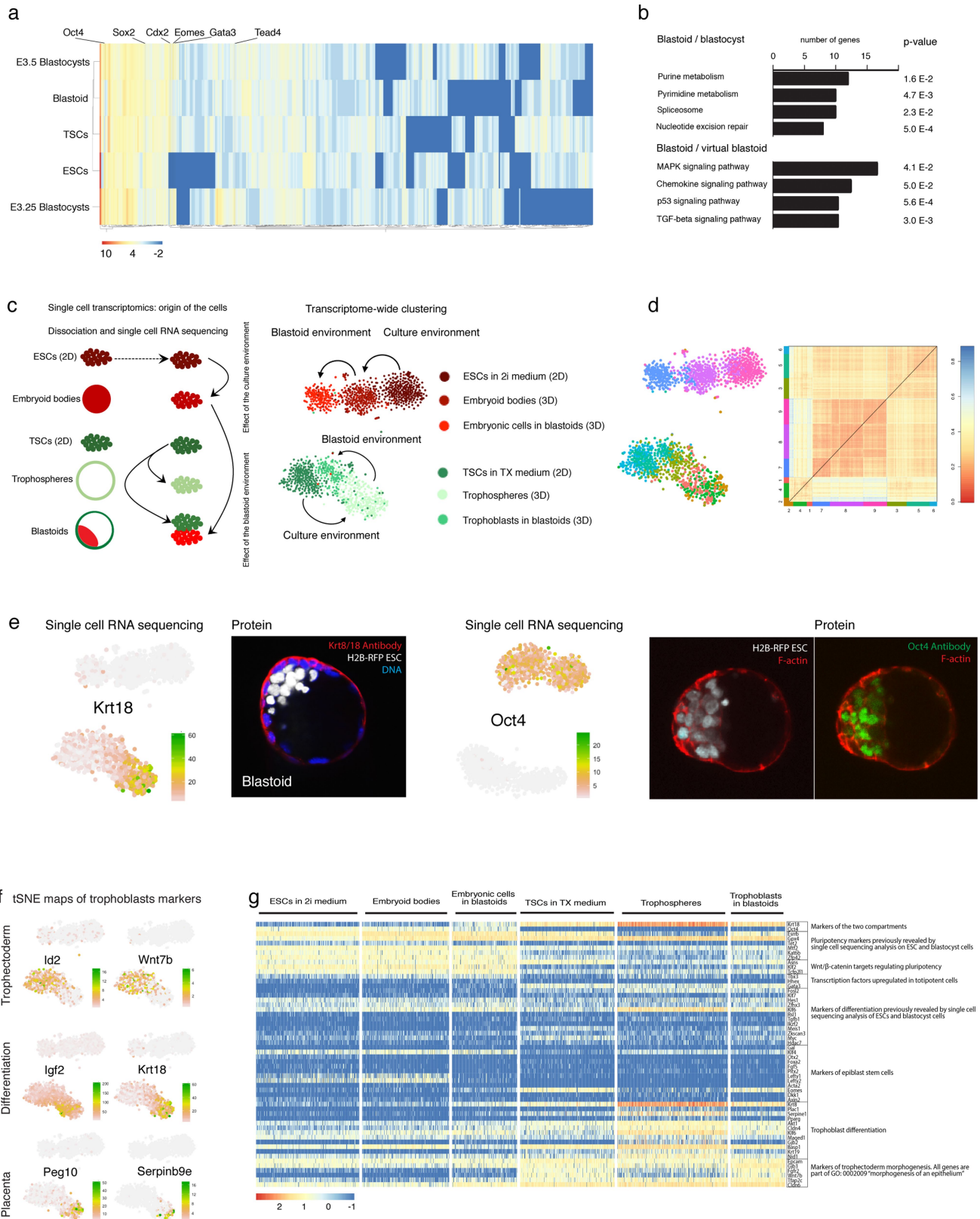
for Nanog, GATA6 and PDGFR α . The number of trophoblast was counted based on the expression of constitutive GFP. The quantification was done at 48, 72 and 86 h. Note that, for most experiments presented herein, blastoids are harvested at 65 h (see Methods). The horizontal bars represent the mean number of cells. The coefficient of variation (relative s.d.) is 19.4% for E3.5 blastocysts and 23.6% for blastoids at 48 h. $n = 6$ independent E3.5 blastocysts and E4.5 blastocysts.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | De novo derivation and developmental potential of blastoid stem cells. **a**, De novo TS and ES cell lines were derived from CDX2-high and CDX2-low blastoids, and blastocysts (see Methods). The derivation was considered successful if colonies appeared on passage 2 after blastocyst or blastoid plating. The derivation efficiency of TS cells from CDX2-high blastoids was significantly higher (**** $P = 0.0001$, Student's t -test) than CDX2-low blastoids and similar to blastocysts (left). $n = 22$ independent blastoids. Error bars are s.d. The potential of de novo-derived cell lines to chimaerize embryos was assessed by immunohistochemistry, using an anti-GFP antibody targeting GFP⁺ TS cells or an anti-RFP antibody targeting H2B-RFP⁺ ES cells (middle). Autofluorescence (denoted by an asterisk), probably due to the presence of blood, occurs within the ecto-placental cone of chimaeric embryos but is lower than the fluorescence recorded from chimaeric embryo after the injection of GFP⁺ TS cells (right). The top image shows a chimaeric embryo after the injection of GFP⁺ TS cells. The bottom image shows a chimaeric embryo after the injection of H2B-RFP⁺ ES cells. Both embryos are imaged in the GFP channel. Note the presence of a

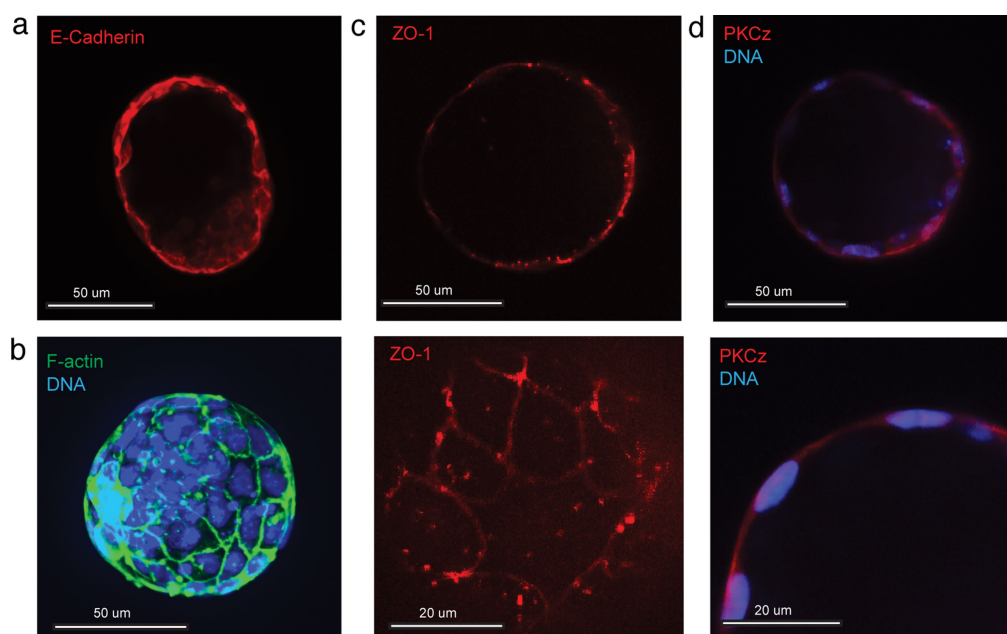
small patch of autofluorescence (denoted by an asterisk) in the bottom image. Microscope settings were adjusted to limit the acquisition of autofluorescence, and images taken upon excitation with a 450–490 nm (green) laser. Scale bars, 100 μm . This experiment is a representation of three similar experiments. **b**, Immunohistochemistry using antibodies against CDX2, ELF5, TEAD4, ASCL2, HAND1 and proliferin, within tissue sections of a representative deciduae including a blastoid grown in utero. Scale bar, 100 μm . This experiment is a representation of three similar experiments. **c**, Percentage of implantation sites in uterus transferred with blastoids and vehicle medium at E3.3–E3.5, and explanted at E7.5. $n = 5$ mice. The value is the mean. Error bars denote the s.d. ** $P = 0.01$, Student's t -test. **d**, Uterus explanted at E7.5 after a physiological development (blastocyst-induced), the transfer of blastoids only in the left horn (blastoid-induced, also see the same uterus in Fig. 2h), and the instillation of oil (oil-induced). **e**, Histological sections of deciduae produced by blastocysts, blastoids and oil instillation, and stained using an anti-ALDH3A1 antibody. Scale bar, 100 μm . This experiment is a representation of three similar experiments.



Extended Data Fig. 6 | See next page for caption.

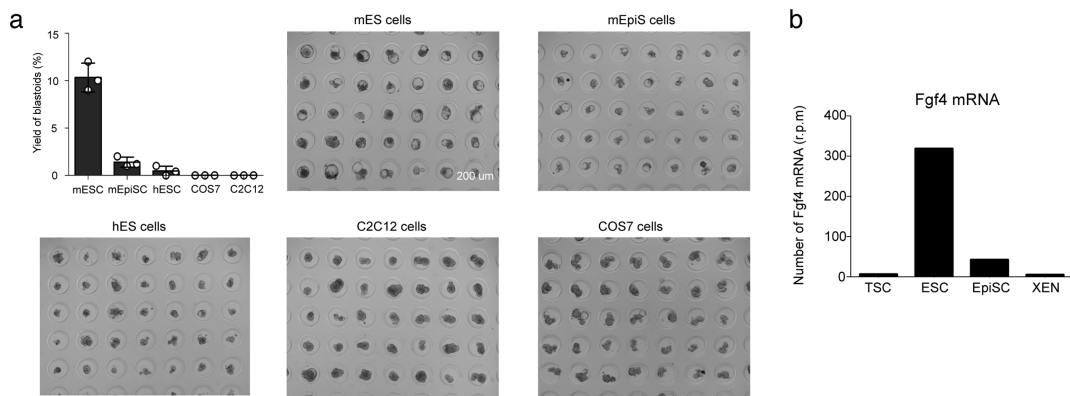
Extended Data Fig. 6 | Bulk and single-cell RNA sequencing of blastoids and blastocysts. **a, b**, Bulk RNA sequencing. **a**, Transcription factor expression distance map. Differentially expressed transcription factors between the ES cells, TS cells, blastoids, virtual blastoids (see Methods), E3.25 and E3.5 blastocysts ($P < 0.05$, DESeq negative binomial distribution) were used to generate a non-supervised distance map of transcription factors selected according to the Riken transcription factor database. The scale is \log_2 . **b**, Gene ontology (GO) and KEGG pathways. Genes differentially regulated ($P < 0.05$, DESeq negative binomial distribution) were analysed using DAVID³⁸ and corresponding GO and KEGG pathways are presented. The full lists of genes related to the depicted GO terms are in Supplementary Table 1 sheet 1. **c–e**, Single-cell sequencing. **c**, Schematic depicting the origin of the single cells. *t*-SNE maps of single cells from ES cells, TS cells, blastoids, trophospheres and embryoid bodies. The colours represent the origin of cells assessed by FACS sorting indexes (Supplementary Table 1 sheet 8). **d**, Clusters of similar cells, as generated by the RACE-ID protocol³⁶. The heat map on the right is the distance map of single cells. The clusters from the *t*-

SNE map and from the heat map are identified using the same colour code. Cells were processed as described in the Methods. **e**, The *Krt18* and *Oct4* genes are markers of the blastocyst trophectoderm and ICM compartments, respectively⁴¹, which we confirmed are also valid to mark the compartments of blastoids using immunohistochemistry (KRT8/18 and OCT4) and H2B-RFP⁺ ES cells. This image is representative of three independent blastoid experiments. **f**, *t*-SNE map representation of key genes for the trophectoderm, trophoblast differentiation and placental cell types. **g**, Heat map showing expression of genes of interest: markers of the trophectoderm and ICM compartments by *Krt18* and *Oct4*, as identified in **e**; markers of pluripotency and differentiation, previously identified in ES cells and blastocyst cells via single-cell sequencing⁴²; genes previously identified as markers of WNT/ β -catenin targets regulating a naive state of pluripotency⁴³; transcription factors upregulated in totipotent cells⁴⁴; genes previously identified as differentiation markers in ES cells via single-cell sequencing⁴²; genes previously identified as markers differentiating ES cells from epiblast stem cells³⁰; markers of trophoblast differentiation; and markers of trophectoderm morphogenesis. The scale is \log_{10} .



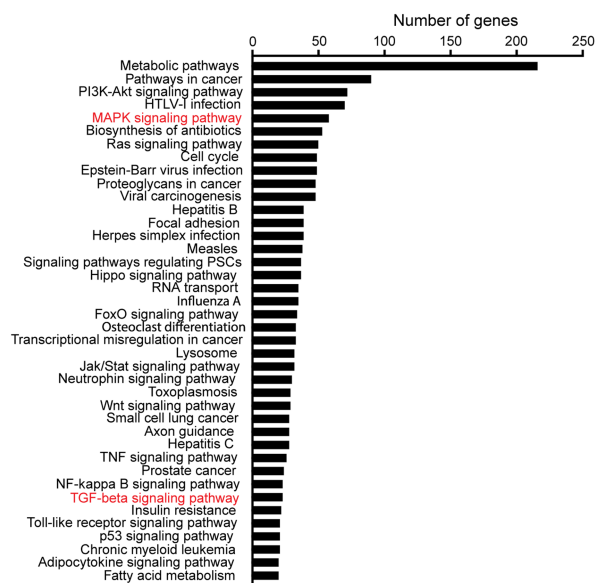
Extended Data Fig. 7 | Immunohistochemistry of epithelial markers within blastoids. **a**, Cross-section of a blastoid stained with an antibody against E-cadherin. Maximum projection of 5 images taken with a 1- μ m step. **b**, Phalloidin staining of the cytoskeletal molecule F-actin and Hoechst staining of DNA show the localization of actin at cell-cell junctions. Maximum projection of 40 images taken with a 1- μ m step.

c, Antibody against the tight junction molecule ZO-1. Section of 0.5 μ m (top). Maximum projection of 4 images taken with a 0.5- μ m step (bottom). **d**, Antibody against the apical molecule PKCz and Hoechst staining of DNA. Section of 1 μ m (top); close up (bottom). All pictures were taken with a PerkinElmer Ultraview VoX spinning disk microscope combined with a Leica SP8. Images are representative of three experiments.



C Inventory of signaling pathways induced in trophoblasts by the embryonic cells.

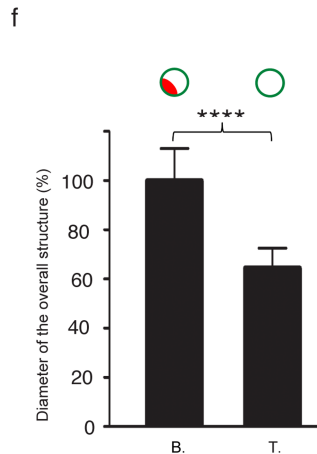
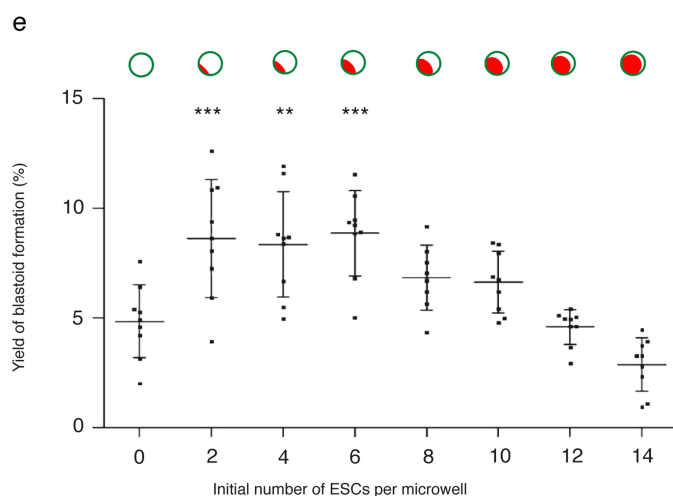
KEGG pathway analysis of trophoblasts from blastoids versus trophospheres



d Genes from selected KEGG pathways

Trophoblasts from blastoids versus trophospheres

Number of genes	MAKING signature	Typh signature	Cell cycle	Local adaptation	Hyposig pathway
1	18	27	49	37	126
2	116	116	140	139	126
3	116	116	140	139	126
4	116	116	140	139	126
5	116	116	140	139	126
6	116	116	140	139	126
7	116	116	140	139	126
8	116	116	140	139	126
9	116	116	140	139	126
10	116	116	140	139	126
11	116	116	140	139	126
12	116	116	140	139	126
13	116	116	140	139	126
14	116	116	140	139	126
15	116	116	140	139	126
16	116	116	140	139	126
17	116	116	140	139	126
18	116	116	140	139	126
19	116	116	140	139	126
20	116	116	140	139	126
21	116	116	140	139	126
22	116	116	140	139	126
23	116	116	140	139	126
24	116	116	140	139	126
25	116	116	140	139	126
26	116	116	140	139	126
27	116	116	140	139	126
28	116	116	140	139	126
29	116	116	140	139	126
30	116	116	140	139	126
31	116	116	140	139	126
32	116	116	140	139	126
33	116	116	140	139	126
34	116	116	140	139	126
35	116	116	140	139	126
36	116	116	140	139	126
37	116	116	140	139	126
38	116	116	140	139	126
39	116	116	140	139	126
40	116	116	140	139	126
41	116	116	140	139	126
42	116	116	140	139	126
43	116	116	140	139	126
44	116	116	140	139	126
45	116	116	140	139	126
46	116	116	140	139	126
47	116	116	140	139	126
48	116	116	140	139	126
49	116	116	140	139	126
50	116	116	140	139	126
51	116	116	140	139	126
52	116	116	140	139	126
53	116	116	140	139	126
54	116	116	140	139	126
55	116	116	140	139	126
56	116	116	140	139	126
57	116	116	140	139	126
58	116	116	140	139	126
59	116	116	140	139	126
60	116	116	140	139	126
61	116	116	140	139	126
62	116	116	140	139	126
63	116	116	140	139	126
64	116	116	140	139	126
65	116	116	140	139	126
66	116	116	140	139	126
67	116	116	140	139	126
68	116	116	140	139	126
69	116	116	140	139	126
70	116	116	140	139	126
71	116	116	140	139	126
72	116	116	140	139	126
73	116	116	140	139	126
74	116	116	140	139	126
75	116	116	140	139	126
76	116	116	140	139	126
77	116	116	140	139	126
78	116	116	140	139	126
79	116	116	140	139	126
80	116	116	140	139	126
81	116	116	140	139	126
82	116	116	140	139	126
83	116	116	140	139	126
84	116	116	140	139	126
85	116	116	140	139	126
86	116	116	140	139	126
87	116	116	140	139	126
88	116	116	140	139	126
89	116	116	140	139	126
90	116	116	140	139	126
91	116	116	140	139	126
92	116	116	140	139	126
93	116	116	140	139	126
94	116	116	140	139	126
95	116	116	140	139	126
96	116	116	140	139	126
97	116	116	140	139	126
98	116	116	140	139	126
99	116	116	140	139	126
100	116	116	140	139	126



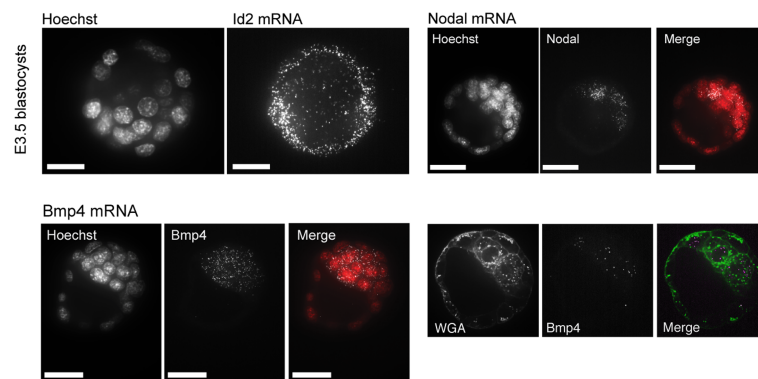
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Assays of embryonic and trophectoderm compartments interactions.

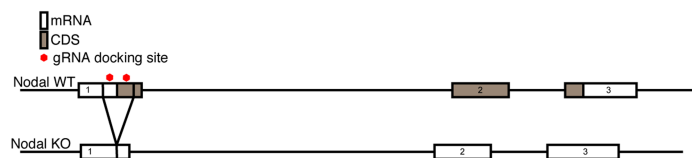
a, Specificity of the ES cell inductions. Yield of formation and representative images of arrays of blastoids resulting from the association of TS cells with ES cells, EpiSCs, human ES cells, COS7 or C2C12 cells. $n = 3$ independent microwell arrays. Centre depicts median values. Error bar denote s.d. **b**, Number of *Fgf4* mRNA transcripts measured by RNA sequencing in TS cells, ES cells, EpiSCs and XEN cells (extra-embryonic endoderm cell lines). **c**, Inventory of signalling pathways induced in trophoblasts by the embryonic cells: KEGG pathways differentiating the trophoblasts from blastoids and trophospheres. The list is exhaustive and generated using the list of statistically differentiated genes ($P < 0.05$, DESeq negative binomial distribution) with the highest fold changes (the 1,500 most highly upregulated and most highly downregulated genes). See the SORT-seq method described previously³³.

d, Selected KEGG pathways and genes related to the MAPK signalling pathway, TGF β signalling pathway, cell cycle, focal adhesion and hippo pathway. **e**, Trophectoderm morphogenesis. Blastoids were assessed based on our definition of cavitated trophoblast structures comprising ES cells, with a circularity greater than 0.9 ($\text{circularity} = 4\pi(\text{area}/\text{perimeter}^2)$), and a diameter between 70 and 110 μm (see Methods). Yield of blastoids (percentage per microwell array) as a function of the initial mean number of ES cells per microwell array. Horizontal bars denote mean yield. Error bars indicate s.d. $**P = 0.01$, $***P = 0.001$, one-way analysis of variance (one-way ANOVA) and Tukey's test. $n = 8$ independent microwell arrays. **f**, The diameter of blastoids and trophospheres measured at 65 h. $n = 50$. $****P = 0.0001$, two-sided Student's t -test. Centre values depict median. Error bar denote s.d. Representative images are shown in Fig. 4a.

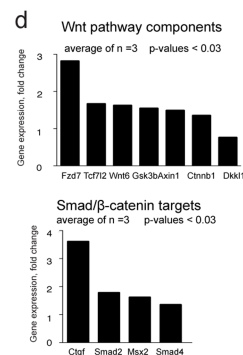
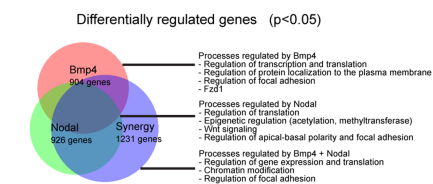
a Single molecule FISH of TGFb pathway components



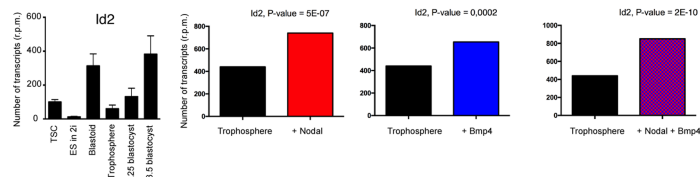
b



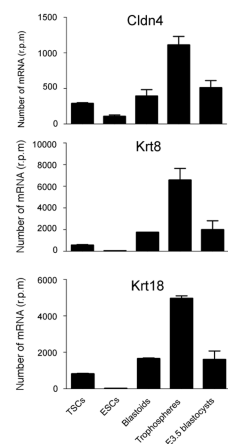
c Differentially regulated genes upon Nodal, Bmp4 or Nodal + Bmp4 stimulation



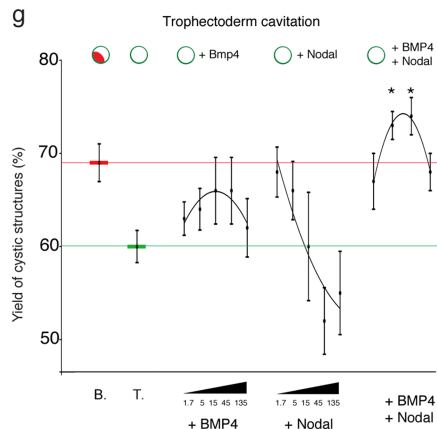
e



f



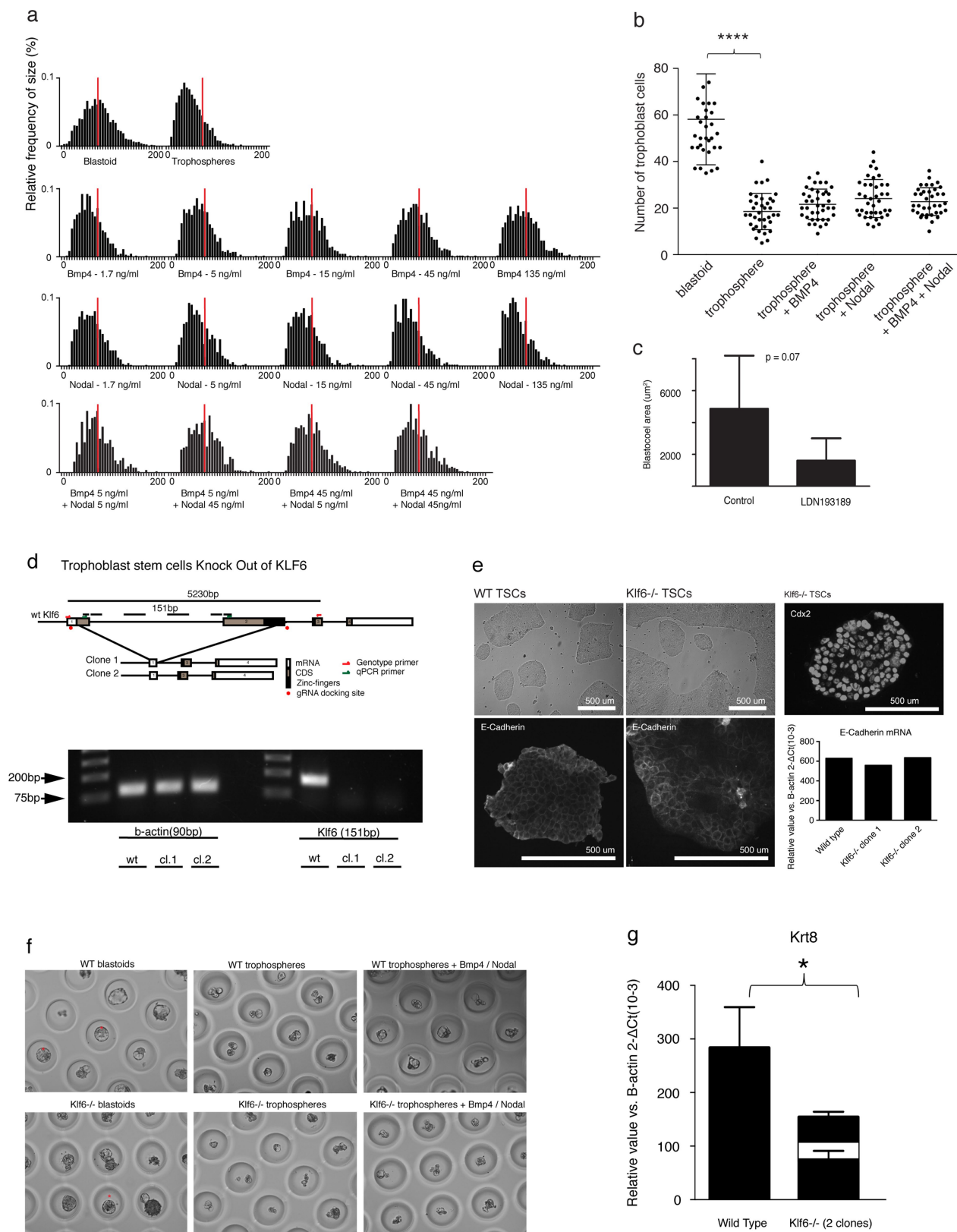
g



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Markers, transcriptional and morphogenetic functions of TGF β signalling. **a**, Single-molecule FISH for *Id2*, *Bmp4* and *Nodal* genes of the TGF β pathway, in an E3.5 blastocyst counterstained with Hoechst. Images are representative of five independent blastocysts. Scales bars are 30 μ m for *Id2* and 50 μ m for *Nodal* and *Bmp4*. **b**, CRISPR strategy for the generation of *Nodal* deletion in ES cells. **c**, Venn diagram of all genes regulated after exposure to BMP4, Nodal or both (synergy). RNA sequencing showed that BMP4 and Nodal regulated a similar number of genes (BMP4: 904 genes, Nodal: 926 genes, genes with $P < 0.05$ (DESeq negative binomial distribution), of which 30% overlapped (413 genes). The GO analysis is shown for each group. See also Supplementary Table 2 sheets 1–4. **d**, RNA-sequencing analysis of the WNT-related genes in trophospheres stimulated with activators of the TGF β signalling pathway. All genes are significantly regulated ($P = 0.03$, DESeq negative binomial distribution). This included the ligand *Wnt6* (1.6-fold, $P = 0.006$, DESeq negative binomial distribution), which is expressed primarily in the cells surrounding the blastocyst cavity¹⁰, the corresponding receptor *Fzd7* (2.8-fold, $P = 0.002$, DESeq negative binomial distribution), the intracellular

effector *Tcf4* (also known as *Tcf7l2*; 1.7-fold, $P = 0.001$, DESeq negative binomial distribution), the negative-feedback regulator *Axin1* (1.5-fold, $P = 0.005$, DESeq negative binomial distribution, top), and of the reported cooperative SMAD/ β -catenin targets *Msx2* (1.6-fold, $P = 0.00003$, DESeq negative binomial distribution)⁴⁵ and *Ctgf* (3.6-fold, $P = 0.003$, DESeq negative binomial distribution)⁴⁶ (bottom). See also Supplementary Table 2 sheets 1–4. **e**, RNA-sequencing analysis of the *Tgfb*-related gene *Id2* in TS cells, ES cells, blastoids, trophospheres, blastocysts (left) and in trophospheres after stimulation with Nodal, BMP4 or both (see Methods) (right). $n = 2$ independent biological samples. The centre depicts the mean. Error bars denote the s.d. **f**, Number of transcripts measured by RNA-sequencing analysis for markers of epithelial development *Cldn4*, *Krt8* and *Krt18*. $n = 2$ independent biological samples. The centre depicts the mean. Error bars denote the s.d. **g**, Yield of cystic structures for a combination of ES and TS cells as compared to TS cells alone (trophospheres) and trophospheres stimulated with BMP4 and Nodal. Horizontal bars denote the mean yield. Error bars are s.d. * $P = 0.02$, Student's *t*-test. $n = 3$ independent microwell arrays (see Methods).



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Generating and assaying blastoids with

***Klf6*^{-/-} TS cells.** **a**, Frequency distribution of the diameter of blastoids, trophospheres and trophospheres exposed to BMP4 and Nodal (μm).

For all graphics, the red bar is the median as measured for blastoids.

b, Number of trophoblasts within blastoids, trophospheres and trophospheres stimulated with activators of the TGF β signalling pathway.

$n = 30$ independent blastoids or trophospheres. $P = 0.0001$, one-way ANOVA. The same graph is shown in part in fig. 4b. **c**, Effect of small-molecule inhibitors on blastocyst cavitation. Morula were flushed from the oviduct of E2.5 CBA \times C57BL/6 mice and cultured in M16 medium. After the initiation of morula compaction, inhibitors or DMSO (1:1,000, control) were added. LDN193189 (0.25 μM) was used to inhibit ALK2 and ALK3 (SMAD pathway). Blastocysts were imaged 24 h after exposure. $n = 10$ independent blastocysts. $P = 0.07$, two-sided Student's t -test. Error bars denote s.d. **d**, Targeting strategy of KLF6 in TS cells and PCR gel of

the obtained clones. **e**, Bright-field image of wild-type and *Klf6*^{-/-} TS cells (top left). Immunostaining for CDX2 in *Klf6*^{-/-} TS cells (top right). E-cadherin immunostaining (bottom left) and quantitative PCR with reverse transcription (qRT-qPCR) (bottom right) of wild-type and *Klf6*^{-/-} TS cells. **f**, Representative pictures of blastoids and trophospheres, and trophospheres stimulated with 45 ng ml⁻¹ BMP4 and 5 ng ml⁻¹ Nodal. Red asterisks denote blastoids, which comply to our definition of cavitated trophoblast structures comprising ES cells, with a circularity greater than 0.9 (circularity = $4\pi(\text{area}/\text{perimeter}^2)$), and a diameter between 70 and 110 μm (see Methods). Comparable results were obtained in three repeated experiments. **g**, qRT-PCR for *Krt8* in structures formed by combining ES cells and wild-type or *Klf6*^{-/-} TS cells. All structures (blastoids and non-blastoids) were collected from the microwell arrays and tested. Horizontal bars indicate mean expression. Error bars denote s.d. $n = 3$ pools from independent microwell arrays. * $P = 0.04$, two-sided Student's t -test.

LUBAC is essential for embryogenesis by preventing cell death and enabling haematopoiesis

Nieves Peltzer^{1,9}, Maurice Darding^{1,9}, Antonella Montinaro¹, Peter Draber^{1,2}, Helena Drabero^{1,2}, Sebastian Kupka¹, Eva Rieser¹, Amanda Fisher³, Ciaran Hutchinson⁴, Lucia Taraborrelli¹, Torsten Hartwig¹, Elodie Lafont¹, Tobias L. Haas⁵, Yutaka Shimizu¹, Charlotta Böiers¹, Aida Sarr¹, James Rickard^{6,7}, Silvia Alvarez-Diaz^{6,7}, Michael T. Ashworth⁴, Allison Beal⁸, Tariq Enver¹, John Bertin⁸, William Kaiser³, Andreas Strasser^{6,7}, John Silke^{6,7}, Philippe Bouillet^{6,7} & Henning Walczak^{1*}

The linear ubiquitin chain assembly complex (LUBAC) is required for optimal gene activation and prevention of cell death upon activation of immune receptors, including TNFR1¹. Deficiency in the LUBAC components SHARPIN or HOIP in mice results in severe inflammation in adulthood or embryonic lethality, respectively, owing to deregulation of TNFR1-mediated cell death^{2–8}. In humans, deficiency in the third LUBAC component HOIL-1 causes autoimmunity and inflammatory disease, similar to HOIP deficiency, whereas HOIL-1 deficiency in mice was reported to cause no overt phenotype^{9–11}. Here we show, by creating HOIL-1-deficient mice, that HOIL-1 is as essential for LUBAC function as HOIP, albeit for different reasons: whereas HOIP is the catalytically active component of LUBAC, HOIL-1 is required for LUBAC assembly, stability and optimal retention in the TNFR1 signalling complex, thereby preventing aberrant cell death. Both HOIL-1 and HOIP prevent embryonic lethality at mid-gestation by interfering with aberrant TNFR1-mediated endothelial cell death, which only partially depends on RIPK1 kinase activity. Co-deletion of caspase-8 with RIPK3 or MLKL prevents cell death in *Hoil-1*^{−/−} (also known as *Rbck1*^{−/−}) embryos, yet only the combined loss of caspase-8 with MLKL results in viable HOIL-1-deficient mice. Notably, triple-knockout *Ripk3*^{−/−} *Casp8*^{−/−} *Hoil-1*^{−/−} embryos die at late gestation owing to haematopoietic defects that are rescued by co-deletion of RIPK1 but not MLKL. Collectively, these results demonstrate that both HOIP and HOIL-1 are essential LUBAC components and are required for embryogenesis by preventing aberrant cell death. Furthermore, they reveal that when LUBAC and caspase-8 are absent, RIPK3 prevents RIPK1 from inducing embryonic lethality by causing defects in fetal haematopoiesis.

To determine the physiological role of HOIL-1, we generated HOIL-1-deficient mice by targeting exons 1 and 2 of the *Hoil-1* (also known as *Rbck1*) gene (Extended Data Fig. 1a–d). No mice with homozygous deletion in the *Hoil-1* gene were weaned (Fig. 1a). Analysis of *Hoil-1*^{−/−} embryos revealed that they died around embryonic day (E) 10.5 (Fig. 1a, b). This result was confirmed with a strain generated from an independently targeted embryonic stem (ES) cell (C20*Hoil-1*^{−/−} mice) (Extended Data Fig. 1e, f). At E10.5, *Hoil-1*^{−/−} embryos presented with disrupted vascular architecture and cell death in the yolk sac endothelium (Fig. 1c, d and Extended Data Fig. 1g, h), indicating that the absence of HOIL-1 causes aberrant endothelial cell death. *Hoil-1*^{fl/fl} *Tie2-cre*⁺ embryos that lack HOIL-1 specifically in endothelial and some haematopoietic cells also died around E10.5 with the same abnormalities (Fig. 1e and Extended Data Fig. 1i, j). Loss of TNF or TNFR1 diminished cell death in the yolk sac and prevented lethality at E10.5 in *Hoil-1*^{−/−} embryos (Fig. 1f and Extended Data Fig. 2a–d). As in the *Tnfr1*^{−/−} *Hoip*^{−/−} (also known as *Tnfrsf1a*^{−/−} *Rnf31*^{−/−}) double

knockouts⁸, *Tnfr1*^{−/−} *Hoil-1*^{−/−} yolk sacs showed reduced cell death as compared to *Hoil-1*^{−/−} embryos (Fig. 1f, g). Although cell death was not completely ablated in *Tnfr1*^{−/−} *Hoil-1*^{−/−} embryos, it did not appear to significantly affect yolk sac vasculature (Fig. 1f, g and Extended Data Fig. 2e). Nevertheless, *Tnfr1*^{−/−} *Hoil-1*^{−/−} embryos died at around E16.5 (Extended Data Fig. 2d, f), with heart defects before death (Fig. 1h). Therefore, like HOIP, HOIL-1 is required to maintain blood vessel integrity by preventing TNFR1-mediated endothelial cell death during embryogenesis.

To understand the role of HOIL-1 in LUBAC function, we compared the formation of the TNFR1 signalling complex (TNFR1-SC) in mouse embryonic fibroblasts (MEFs) individually deficient for the LUBAC components. Although TNFR1-SC-associated linear ubiquitination was merely reduced in SHARPIN-deficient MEFs⁷, it was completely absent in *Tnfr1*^{−/−} *Hoil-1*^{−/−} MEFs, exactly as in *Tnfr1*^{−/−} *Hoip*^{−/−} MEFs⁸ (Fig. 2a). In TNF-stimulated *Tnfr1*^{−/−} *Hoil-1*^{−/−} MEFs, NF-κB activation was attenuated (Extended Data Fig. 3a) and TNFR1 complex-II formation was enhanced (Fig. 2b), resulting in sensitization to TNF-induced apoptosis and necroptosis (Fig. 2c). Hence, HOIL-1 is as essential as HOIP for linear ubiquitination within the TNFR1-SC.

To determine whether the reduction in HOIP and SHARPIN protein levels in HOIL-1-deficient cells was responsible for the observed loss of linear ubiquitination (Fig. 2a), we reconstituted HOIL-1-deficient MEFs with HOIP, with HOIP plus SHARPIN, or, as a control, with HOIL-1. Reconstitution with HOIP, either alone or with SHARPIN, failed to restore LUBAC recruitment, linear ubiquitination at the TNFR1-SC, or optimal NF-κB activation. Furthermore, the reconstitution of HOIP and/or SHARPIN was unable to prevent TNF-induced complex-II formation and cell death, whereas the re-expression of HOIL-1 corrected all aforementioned defects (Fig. 2d–f and Extended Data Fig. 3b). In the absence of HOIL-1, HOIP was unable to bind to SHARPIN despite both being reconstituted to near endogenous levels (Extended Data Fig. 3c). Thus, HOIL-1 is required for LUBAC assembly and recruitment to the TNFR1-SC, identifying it as an essential component of LUBAC alongside HOIP.

To reveal how HOIL-1 enables LUBAC activity, we generated HOIL-1-deficient MEFs stably expressing full-length wild-type HOIL-1, the UBL domain of HOIL-1 only (HOIL-1-UBL), HOIL-1-ΔRBR, HOIL-1-ΔUBL, HOIL-1 with inactivating mutations T201A/R208A in the NZF domain (HOIL-1-NZFmut) or HOIL-1 with a point mutation in the catalytic cysteine of the RBR domain (HOIL-1-C458A) (Fig. 2g). Except for HOIL-1-ΔUBL, all mutant HOIL-1 proteins bound to HOIP and SHARPIN and stabilized their levels (Fig. 2h). Isolation of the native TNFR1-SC revealed that HOIL-1-ΔRBR and HOIL-1-C458A fully restored TNF-induced linear ubiquitination in HOIL-1-deficient cells, whereas HOIL-1-ΔUBL did not (Fig. 2i). HOIL-1-deficient cells

¹UCL Cancer Institute, University College London, London, UK. ²Laboratory of Adaptive Immunity, Institute of Molecular Genetics, Czech Academy of Sciences, Prague, Czech Republic. ³University of Texas Health Science Center, San Antonio, TX, USA. ⁴UCL Great Ormond Street Institute of Child Health, London, UK. ⁵Institute of General Pathology, Università Cattolica del Sacro Cuore, Rome, Italy. ⁶The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. ⁷Department of Medical Biology, The University of Melbourne, Melbourne, Victoria, Australia. ⁸Pattern Recognition Receptor Discovery Performance Unit, Immuno-Inflammation Therapeutic Area, GlaxoSmithKline, Collegeville, PA, USA. ⁹These authors contributed equally: Nieves Peltzer, Maurice Darding. *e-mail: h.walczak@ucl.ac.uk

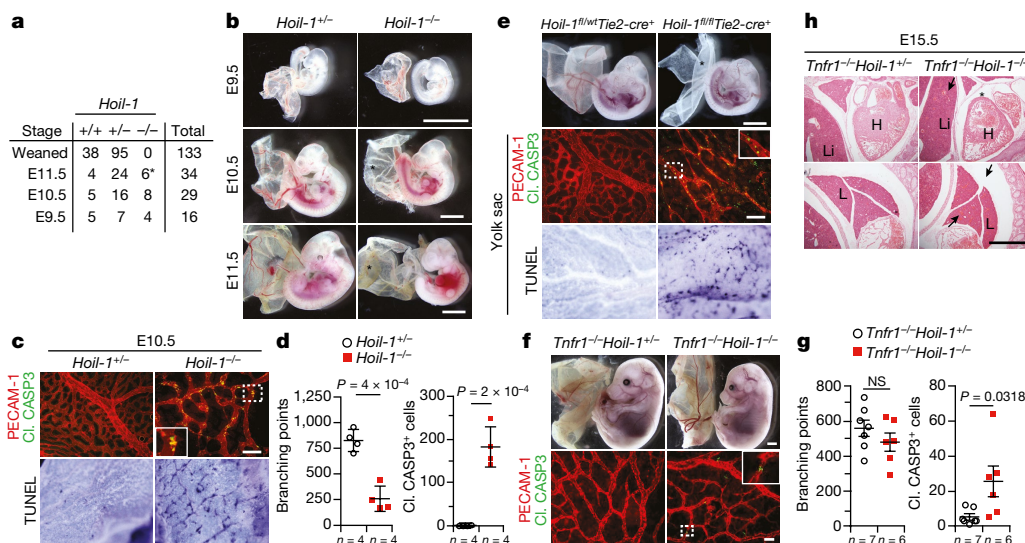


Fig. 1 | HOIL-1 deficiency causes embryonic lethality at mid-gestation due to TNFR1-mediated endothelial cell death. **a**, Mendelian frequencies obtained from inter-crossing *Hoil-1*^{+/+} mice. Asterisk denotes dead embryos. **b**, Representative images of embryos from E9.5 to E11.5 quantified in **a**. Asterisks denote poor yolk sac vascularization. Scale bars, 2 mm. **c**, Top, representative images of yolk sac vascularization (PECAM-1, red) and cell death (cleaved (CI) CASP3 staining, green) at E10.5 ($n = 4$ yolk sacs per genotype). Bottom, whole-mount TUNEL staining ($n = 2$ yolk sacs per genotype). Scale bars, 50 μ m. **d**, **g**, Quantification of branching points and cleaved CASP3-positive cells in **c** and **f**. Data are mean \pm s.e.m. P values from unpaired two-tailed t -tests are shown. NS, not significant. **e**, Top, representative images of embryos at E10.5 ($n = 14$ *Hoil-1*^{fl/wt}*Tie2-cre*⁺ and $n = 7$ *Hoil-1*^{fl/fl}*Tie2-cre*⁺ embryos, top). Asterisk

denotes poor yolk sac vascularization. Scale bar, 2 mm. Middle, yolk sac vascularization (PECAM-1, red) and apoptosis (cleaved CASP3, green). Scale bar, 50 μ m. Bottom, yolk sac whole-mount TUNEL staining ($n = 6$ *Hoil-1*^{fl/wt}*Tie2-cre*⁺ and $n = 2$ *Hoil-1*^{fl/fl}*Tie2-cre*⁺ yolk sacs per genotype). **f**, Top, representative images of embryos at E15.5 ($n = 6$ *Tnfr1*^{-/-}*Hoil-1*^{+/+} and $n = 19$ *Tnfr1*^{-/-}*Hoil-1*^{-/-} embryos). Scale bar, 2 mm. Bottom, yolk sac vascularization (PECAM-1, red) and apoptosis (cleaved CASP3, green). Scale bar, 50 μ m. **h**, Representative images of haematoxylin and eosin (H&E) staining on whole-embryo paraffin sections ($n = 3$ embryos per genotype). Asterisk denote pericardial effusion, arrows denote congested vessels. H, heart; L, lung; Li, liver. Scale bar, 50 μ m.

expressing HOIL-1-UBL or HOIL-1-NZFmut only showed partial restoration of linear ubiquitination, correlating with reduced HOIP and SHARPIN levels at the TNFR1-SC (Fig. 2i). Thus, the UBL domain of HOIL-1 is essential for linear ubiquitination at the TNFR1-SC, whereas a functional NZF domain is required for optimal LUBAC presence in the TNFR1-SC. Expression of HOIL-1- Δ RBR restored optimal NF- κ B signalling and prevented aberrant TNF-induced cell killing in contrast to HOIL-1- Δ UBL (Fig. 2j and Extended Data Fig. 3d). This observation explains why the previously reported mice, regarded as deficient for HOIL-1, are viable as they were generated by targeting exons 7 and 8¹¹, probably resembling the HOIL-1- Δ RBR mutant studied here. Because the UBL of HOIL-1 binds to HOIP, allowing its activation¹², and the NZF of HOIL-1 binds linear ubiquitin linkages¹³ our results provide evidence that HOIL-1 promotes HOIP activation as well as LUBAC assembly and recruitment to the TNFR1-SC via its UBL domain. Once linear ubiquitin chains are formed in the complex, the NZF domain of HOIL-1 promotes LUBAC retention by binding to these chains.

Because both HOIL-1 and HOIP are equally important for LUBAC function and, consequently, for preventing aberrant cell death in vitro and in vivo, we used a genetic strategy to untangle the interaction between HOIL-1 or HOIP and the different cell death components. Inactivation of RIPK1 in *Hoil-1*^{-/-} and *Hoip*^{-/-} embryos delayed lethality until E14.5 (Fig. 3a and Extended Data Fig. 4a–d). At this time, *Ripk1*^{K45A}*Hoil-1*^{-/-} and *Ripk1*^{K45A}*Hoip*^{-/-} embryos had disrupted vascular architecture, excessive cell death in their yolk sacs, hearts, livers and lungs, and presented with heart defects and liver necrosis (Fig. 3b and Extended Data Fig. 4e–h). In accordance, TNFR1 complex-II formation and aberrant apoptosis induced by TNF or lymphotoxin- α (LT- α) were only partially inhibited in RIP1 kinase-dead *Ripk1*^{K45A}*Hoil-1*^{-/-} MEFs (Fig. 3c, d and Extended Data Fig. 4i). Thus, although the kinase activity of RIPK1 is essential for excessive TNFR1-induced cell death caused by attenuated LUBAC activity, as previously observed in SHARPIN-deficient mice⁴, this is not the case when LUBAC activity is completely abrogated.

We next tested whether the loss of RIPK3, MLKL or caspase-8 could prevent lethality in *Hoip*^{-/-} and *Hoil-1*^{-/-} embryos. At E10.5, *Ripk3*^{-/-}*Hoil-1*^{-/-} embryos presented with defects in vascularization, excessive cell death and died at mid-gestation (Extended Data Fig. 5b, c). Owing to the close chromosomal linkage of HOIP and RIPK3, we generated *Mkl1*^{-/-}*Hoip*^{-/-} mice (Extended Data Fig. 5a). These embryos also died at mid-gestation (Extended Data Fig. 5d). Likewise, neither *Casp8* heterozygosity nor full deletion was sufficient to prevent the mid-gestation lethality of *Hoip*^{-/-} and *Hoil-1*^{-/-} embryos (Extended Data Fig. 5e, f and data not shown).

As RIPK3-mediated necroptosis may be responsible for the embryonic lethality of *Casp8*^{+/+}*Hoil-1*^{-/-} or *Casp8*^{-/-}*Hoil-1*^{-/-} mice^{14,15}, we generated *Ripk3*^{-/-}*Casp8*^{+/+}*Hoil-1*^{-/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos and in both cases the lethality was delayed until around E14.5 (Fig. 3e and Extended Data Fig. 6a, b). At this developmental stage, a single intact copy of caspase-8 was sufficient to induce apoptosis-driven loss of yolk sac vascularization (Fig. 3f and Extended Data Fig. 6c, d). Yet, although *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos died around E14.5, yolk sac vascularization was normalized and cell death in the yolk sac and other organs was prevented (Fig. 3f and Extended Data Fig. 6c–f). Moreover, *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} MEFs were resistant to cell death induced by TNF or related cytokines (Extended Data Fig. 6g). Histological examination and microfocus computed tomography scanning revealed the presence of heart defects in both *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and *Ripk3*^{-/-}*Casp8*^{+/+}*Hoil-1*^{-/-} embryos (Extended Data Fig. 6h, i). We therefore conclude that whereas mid-gestation lethality in *Hoil-1*^{-/-} embryos is dependent on caspase-8/RIPK3-mediated apoptosis and necroptosis, *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos die at late gestation by a process that is independent of cell death.

In marked contrast to *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} mice, both *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and *Mkl1*^{-/-}*Casp8*^{-/-}*Hoip*^{-/-} mice were born, albeit at lower than expected Mendelian ratios (Fig. 3g and Extended Data Fig. 7a). These mice were runted and had to be

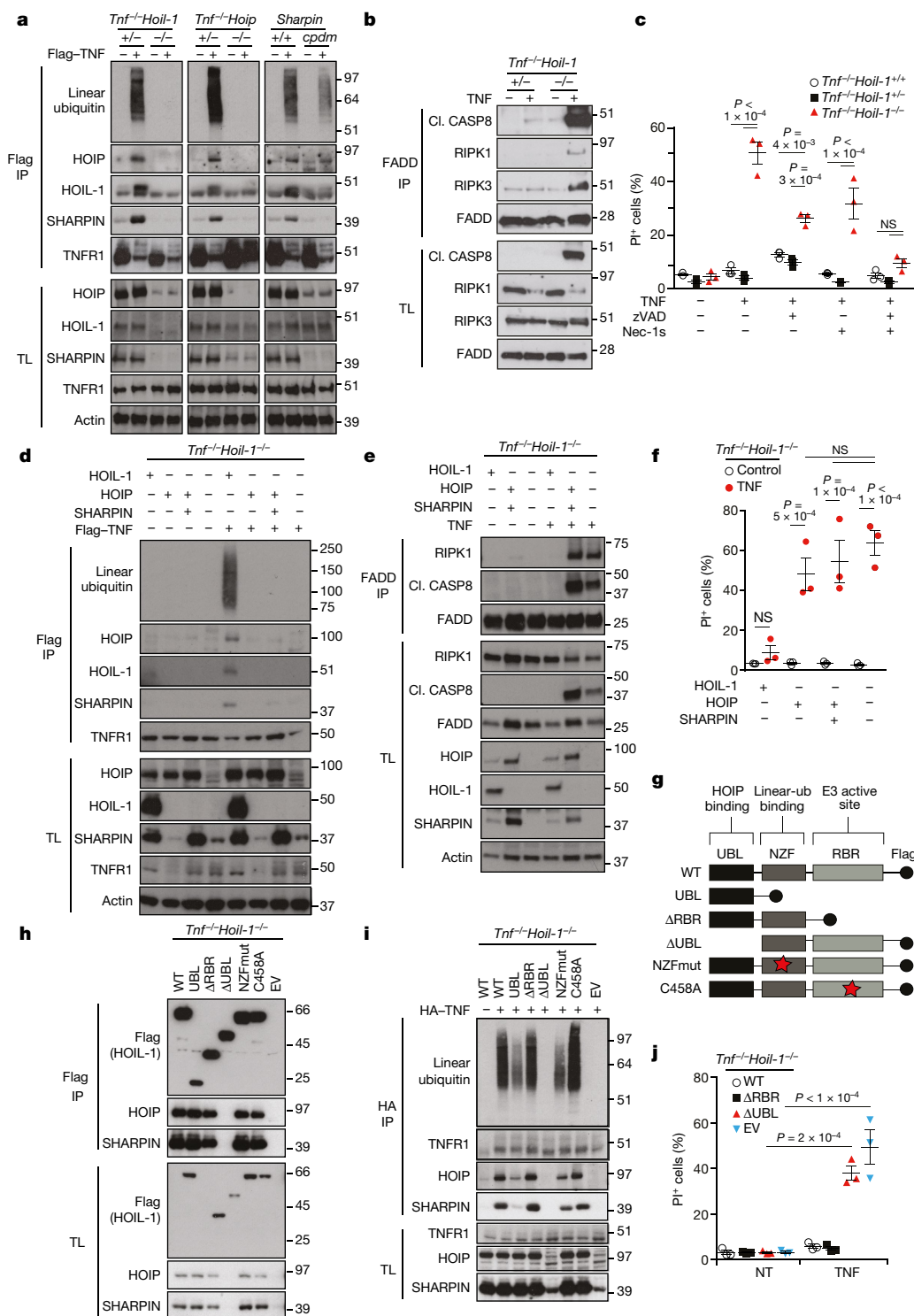


Fig. 2 | The UBL domain but not the RBR domain of HOIL-1 is essential for LUBAC activity at the TNFR1-SC and to prevent TNF/TNFR1-induced cell death. **a, d,** TNFR1-SC pull-down by Flag immunoprecipitation (IP) in MEFs derived from mice of the indicated genotypes \pm Flag-TNF for 15 min ($n = 2$ independent experiments) (**a**) and reconstituted with HOIL-1, HOIP or HOIP and SHARPIN ($n = 4$ independent experiments) (**d**). **b, e,** Immunoprecipitation of the adaptor protein FADD in MEFs of the indicated genotypes treated for 4 h with the caspase inhibitor zVAD-fmk \pm TNF (**b**) and reconstituted as indicated (**e**) ($n = 2$ independent experiments (**b, e**)). **c, f, j,** Cell death analysed by propidium iodide (PI) staining in MEFs with the indicated

genotypes \pm TNF \pm the indicated inhibitors zVAD-fmk (zVAD) and/or necrostatin-1 (Nec-1s) for 24 h (**c**), reconstituted (**f**) or transduced (**j**) as indicated. Mean \pm s.e.m. ($n = 3$ independent experiments) and P values from two-way ANOVA are shown. **g,** Schematic overview of HOIL-1 constructs used to transduce *Tnfr^{-/-}Hoil-1^{-/-}* MEFs. WT, wild type. **h,** Flag immunoprecipitation of indicated HOIL-1 mutants ($n = 2$ independent experiments). **i,** Endogenous TNFR1-SC pull-down by haemagglutinin (HA) immunoprecipitation in reconstituted *Tnfr^{-/-}Hoil-1^{-/-}* MEFs \pm HA-TNF for 15 min ($n = 2$ independent experiments). EV, empty vector; NT, not treated; TL, total lysate. For gel source data (**a, b, d, e, h, i**), see Supplementary Fig. 1.

ethanized by 4–5 weeks of age (Fig. 3h, Extended Data Fig. 7b, c). Histopathological analysis revealed severe inflammation in the liver and lungs (Extended Data Fig. 7d and data not shown). Of note,

Casp8 heterozygosity resulted in increased apoptosis of endothelial cells, causing lethality in both *Mkl1^{-/-}Casp8^{+/-}Hoip^{-/-}* and *Mkl1^{-/-}Casp8^{+/-}Hoil-1^{-/-}* embryos around E14.5 (Extended Data

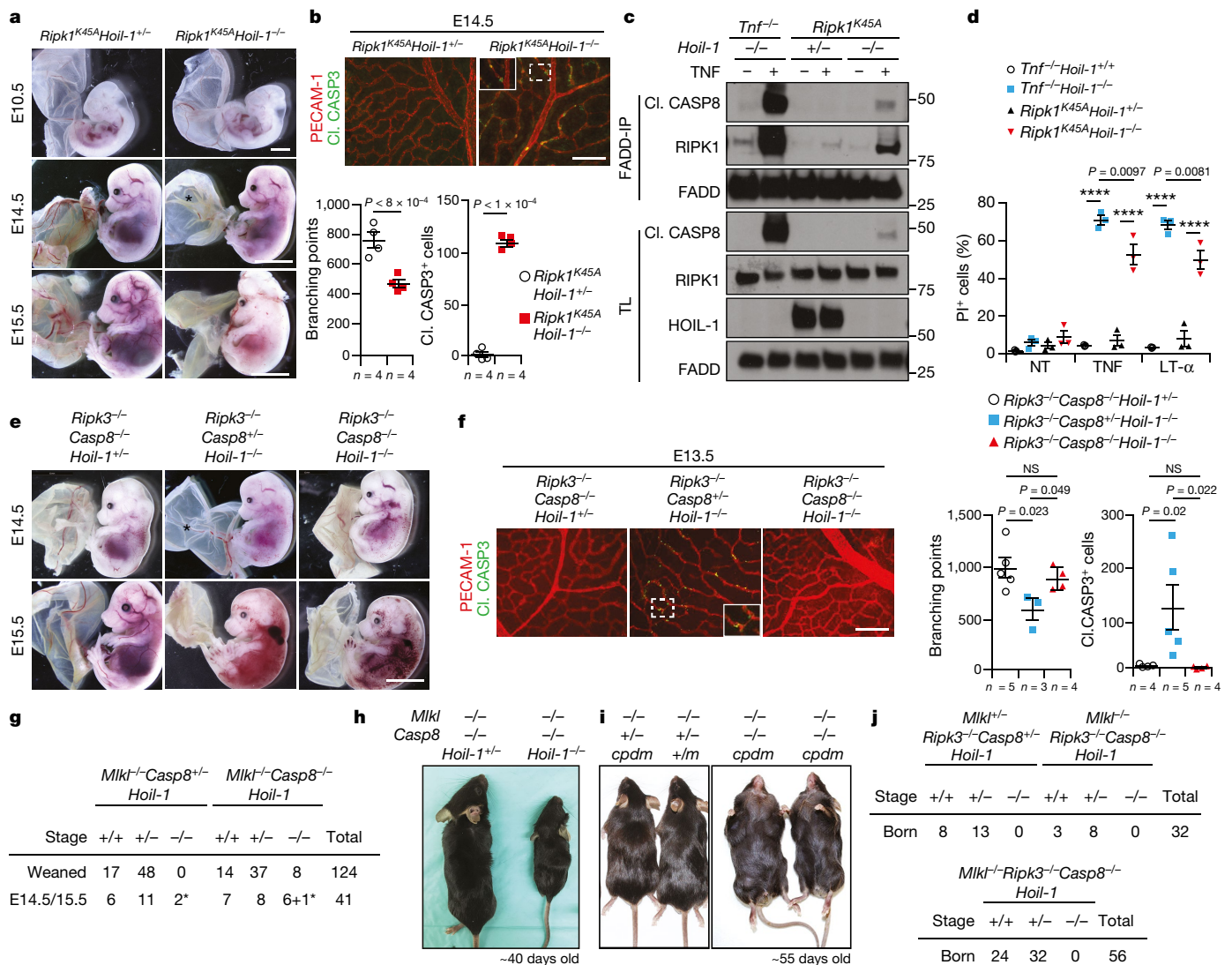


Fig. 3 | Concomitant loss of MLKL and caspase-8, but not loss of RIPK1 kinase activity or combined loss of RIPK3 and caspase-8, promotes survival of LUBAC-deficient mice. **a**, Representative images of E10.5 ($n = 6$ embryos per genotype), E14.5 ($n = 12$ *Ripk1^{K45A}Hoil-1^{+/-}*, $n = 5$ *Ripk1^{K45A}Hoil-1^{-/-}* embryos per genotype) and E15.5 embryos ($n = 3$ embryos per genotype). Scale bars, 2 mm (E10.5) and 5 mm (E14.5, E15.5). Asterisk denotes poor yolk sac vascularization. **b**, **f**, Representative images of yolk sac vascularization (PECAM-1, red) and apoptosis (cleaved (CL) CASP3, green) at E14.5 (**b**) or E13.5 (**f**) and quantification. Mean \pm s.e.m. and P values from unpaired two-tailed t -tests (**b**) or one-way ANOVA (**f**) are shown. Scale bar, 50 μ m. **c**, Immunoprecipitation of the adaptor protein FADD in MEFs treated for 3 h with TNF and zVAD-fmk ($n = 2$ independent experiments). For gel source data, see Supplementary Fig. 1. **d**, Cell death measured by propidium iodide (PI)

Fig. 7e and data not shown) indicating that caspase-8-driven apoptosis is sufficient to cause embryonic death of LUBAC-deficient embryos.

Co-deletion of RIPK3 and caspase-8 causes embryonic lethality in otherwise viable SHARPIN-deficient *cpdm* (chronic proliferative dermatitis mice, also known as *Sharpin^{cpdm}*) mice⁷. However, *Mkl^{-/-}Casp8^{-/-}Sharpin^{cpdm}* mice were viable and the inflammatory syndrome that characterizes *Sharpin^{cpdm}* mice was prevented (Fig. 3i and Extended Data Fig. 7f, g), while expected¹⁶ developing lymphadenopathy and splenomegaly (Fig. 3i and Extended Data Fig. 7f). Thus, the combined loss of any of the three LUBAC components together with the loss of caspase-8 uncovers a vital functional difference between RIPK3 and MLKL.

We next evaluated whether the lethality of *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* mice is due to aberrant (RIPK3-independent) MLKL

incorporation in MEFs treated with TNF (10 ng ml⁻¹) or LT- α , or not treated (NT). Data are mean \pm s.e.m. ($n = 3$ independent experiments). **** $P < 0.0001$, two-way ANOVA. **e**, Representative images of E14.5 ($n = 11$ *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{+/-}*, *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* and $n = 7$ *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}*) and E15.5 ($n = 5$ *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{+/-}*, $n = 4$ *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* and $n = 8$ *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}*) embryos. Asterisk denotes poor yolk sac vascularization. Scale bar, 5 mm. **g**, **j**, Mendelian frequencies obtained from intercrossing *Mkl^{-/-}Casp8^{-/-}Hoil-1^{+/-}* with *Mkl^{-/-}Casp8^{-/-}Hoil-1^{+/-}* mice (**g**) or *Mkl^{-/-}Ripk3^{-/-}Casp8^{-/-}Hoil-1^{+/-}* with *Mkl^{-/-}Ripk3^{-/-}Casp8^{-/-}Hoil-1^{+/-}* mice (**j**, top) or *Mkl^{-/-}Ripk3^{-/-}Casp8^{-/-}Hoil-1^{+/-}* mice (**j**, bottom). Asterisk denotes dead embryo. **h**, **i**, Representative images of adult mice quantified in **g** (**h**), or $n = 3$ mice per genotype in **i**. *m* denotes *cpdm* mutation.

activation. This was particularly pertinent because MLKL levels were increased in *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* embryos and MLKL was aberrantly activated in some of them (Extended Data Fig. 7h). However, MLKL co-deficiency did not prevent the death of *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* embryos (Fig. 3j). Thus, RIPK3 is required for the survival of embryos in the absence of LUBAC by regulating an MLKL-independent process.

To explore the nature of the pro-survival role of RIPK3, we performed RNA sequencing (RNA-seq) on E13.5 *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}*, *Mkl^{-/-}Casp8^{-/-}Hoil-1^{-/-}* and control embryos (Extended Data Fig. 8a and Supplementary Table 1). Gene Ontology (GO) enrichment analysis of differentially expressed genes pointed towards defects in erythropoiesis in *Ripk3^{-/-}Casp8^{-/-}Hoil-1^{-/-}* embryos (Extended Data Fig. 8b). Indeed, reduced levels of erythroid lineage TER119⁺ cells

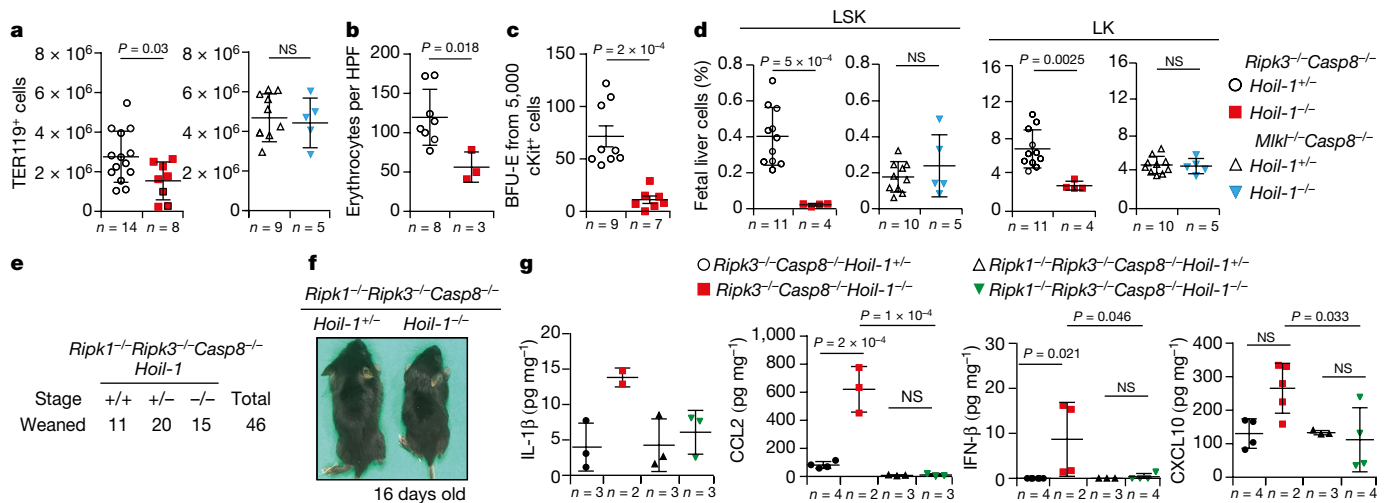


Fig. 4 | Combined deletion of RIPK3 and caspase-8 causes haematopoietic defects and RIPK1-dependent embryonic lethality in HOIL-1-deficient mice. **a**, **b**, Number of TER119⁺ (erythroid) cells (**a**) and enucleated erythrocytes per high-power field (HPF) (**b**) in E13.5 fetal livers with the indicated genotypes. Data are mean \pm s.e.m. P values from unpaired two-tailed t -tests are shown. **c**, Differentiation of E13.5 fetal liver (c-KIT⁺) progenitors into erythroid burst-forming units (BFU-E). Data are mean \pm s.e.m. P values from unpaired two-tailed t -tests are reported. **d**, Percentage of haematopoietic progenitors negative for

mature lineage markers (Lin⁻) and SCA-1⁺c-KIT⁺ (LSK) and SCA-1⁻c-KIT⁺ (LK) in E13.5 fetal livers with the indicated genotypes. Data are mean \pm s.e.m. P values from unpaired two-tailed t -tests are reported. **e**, Mendelian frequencies obtained from intercrossing *Ripk1*^{-/-}*Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} mice. **f**, Representative images of mice of the indicated genotypes quantified in **e**. **g**, Cytokine levels in embryo homogenates with the indicated genotypes. Data are mean \pm s.e.m. P values from one-way ANOVA are reported.

(Fig. 4a), basophilic erythroblasts (Extended Data Fig. 8c) and mature erythrocytes (Fig. 4b) were observed in *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal livers. Furthermore, *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} haematopoietic progenitors failed to differentiate into committed erythroid burst-forming units (BFU-E) in culture (Fig. 4c). Further analysis of the haematopoietic compartment from E13.5 fetal livers revealed abnormally reduced percentages and total numbers of multipotent progenitors (Fig. 4d and Extended Data Fig. 8d, e) as well as leucocytes, including granulocytes and macrophages, and myeloid progenitors in the *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos compared to controls, whereas *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos had normal numbers of these cells (Extended Data Fig. 8f–k). In addition, the capacity of haematopoietic progenitors to generate colony-forming myeloid progenitors and multipotent progenitors was also impaired in the *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos (Extended Data Fig. 8l). Accordingly, the viability of macrophages obtained from *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal liver cell suspensions in culture was significantly lower than those of controls and this could not be rescued by inhibiting necroptosis or apoptosis. *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal liver cells, however, produced normal numbers of macrophages (Extended Data Fig. 4m). Despite the heart defects of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos, blood circulation was normal at E13.5 and the percentages of CD45⁺cKIT⁺ cells obtained from aorta–gonad–mesonephros (AGM) regions were comparable between *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos and controls at E11.5 (Extended Data Fig. 8o, p). We therefore conclude that *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos have defective early haematopoiesis, probably downstream of specification in the AGM, resulting in substantial deficiencies in erythroid and myeloid cells.

Because LUBAC is known to regulate RIPK1^{17,18}, we investigated the role of RIPK1 in the lethality of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos. The lethality of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos was prevented by additional loss of RIPK1, despite RIPK1 levels being relatively low in *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos and RIPK1 deficiency failing to prevent *Hoil-1*^{-/-} embryonic lethality (Fig. 4d, e and Extended Data Figs. 7h, 9a, b). Importantly, the viability of macrophages obtained from *Ripk1*^{-/-}*Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal livers was comparable to controls (Extended Data Fig. 9c), indicating normalized haematopoiesis in these mice. The expression of several cytokines, including IL-1 β , CCL2, IFN- β and CXCL10,

was abnormally increased in *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos but not in *Ripk1*^{-/-}*Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos (Fig. 4f and Extended Data Fig. 9d, e). The function, survival, differentiation and self-renewal of haematopoietic progenitors are greatly impacted by several of these cytokines^{19,20}. Therefore, our findings suggest that RIPK1-driven deregulated cytokine production in *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos may impair fetal haematopoiesis. Finally, the treatment of pregnant females with the RIPK1 kinase inhibitor GSK3540547A (GSK547A)²¹ did not prevent lethality of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos, although it was able to extend the survival of *Ripk3*^{-/-}*Casp8*^{+/-}*Hoil-1*^{-/-} embryos (Extended Data Fig. 9f). These results suggest that the lethality of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos probably depends on the scaffolding function of RIPK1.

Although RIPK1 is required for emergency haematopoiesis, RIPK1 might regulate embryonic haematopoiesis differently. Indeed, RIPK1-constitutive or RIPK1-haematopoietic-cell-specific-deficient mice are not embryonically lethal^{22,23}. In addition, the absence of LUBAC, RIPK3 and caspase-8 might affect mechanisms during embryogenesis that are different from those perturbed by RIPK1 deficiency alone. Collectively, our findings indicate that in the combined absence of LUBAC and caspase-8, RIPK3 exerts a pro-survival role by regulating RIPK1-mediated signalling (Extended Data Fig. 10). Because *Ripk3*^{-/-}*Casp8*^{-/-} mice are viable^{14,15,24}, our findings indicate that the control of RIPK1 by either LUBAC or RIPK3 is sufficient to enable proper haematopoiesis in the developing embryo, probably by preventing deregulated cytokine production. Thus, LUBAC and RIPK3 control RIPK1-mediated signalling to allow embryonic haematopoiesis.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0064-8>.

Received: 21 October 2016; Accepted: 16 March 2018;

Published online 25 April 2018.

- Shimizu, Y., Taraborrelli, L. & Walczak, H. Linear ubiquitination in immunity. *Immunol. Rev.* **266**, 190–207 (2015).

2. Sasaki, Y. et al. Defective immune responses in mice lacking LUBAC-mediated linear ubiquitination in B cells. *EMBO J* **32**, 2463–2476 (2013).
 3. Emmerich, C. H. et al. Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *Proc. Natl Acad. Sci. USA* **110**, 15247–15252 (2013).
 4. Berger, S. B. et al. Cutting edge: RIP1 kinase activity is dispensable for normal development but is a key regulator of inflammation in SHARPIN-deficient mice. *J. Immunol.* **192**, 5476–5480 (2014).
 5. Gerlach, B. et al. Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596 (2011).
 6. Kumari, S. et al. Sharpin prevents skin inflammation by inhibiting TNFR1-induced keratinocyte apoptosis. *eLife* **3**, e03422 (2014).
 7. Rickard, J. A. et al. TNFR1-dependent cell death drives inflammation in Sharpin-deficient mice. *eLife* **3**, e03464 (2014).
 8. Peltzer, N. et al. HOIP deficiency causes embryonic lethality by aberrant TNFR1-mediated endothelial cell death. *Cell Reports* **9**, 153–165 (2014).
 9. Boisson, B. et al. Immunodeficiency, autoinflammation and amylopectinosis in humans with inherited HOIL-1 and LUBAC deficiency. *Nat. Immunol.* **13**, 1178–1186 (2012).
 10. Boisson, B. et al. Human HOIP and LUBAC deficiency underlies autoinflammation, immunodeficiency, amylopectinosis, and lymphangiectasia. *J. Exp. Med.* **212**, 939–951 (2015).
 11. Tokunaga, F. et al. Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nat. Cell Biol.* **11**, 123–132 (2009).
 12. Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E. & Rittinger, K. LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO Rep.* **13**, 840–846 (2012).
 13. Sato, Y. et al. Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. *Proc. Natl Acad. Sci. USA* **108**, 20520–20525 (2011).
 14. Kaiser, W. J. et al. RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* **471**, 368–372 (2011).
 15. Oberst, A. et al. Catalytic activity of the caspase-8-FLIP_L complex inhibits RIPK3-dependent necrosis. *Nature* **471**, 363–367 (2011).
 16. Alvarez-Diaz, S. et al. The pseudokinase MLKL and the kinase RIPK3 have distinct roles in autoimmune disease caused by loss of death-receptor-induced apoptosis. *Immunity* **45**, 513–526 (2016).
 17. Peltzer, N., Darding, M. & Walczak, H. Holding RIPK1 on the ubiquitin leash in TNFR1 signaling. *Trends Cell Biol.* **26**, 445–461 (2016).
 18. Annibaldi, A. & Meier, P. Checkpoints in TNF-induced cell death: implications in inflammation and cancer. *Trends Mol. Med.* **24**, 49–65 (2018).
 19. Clapes, T., Lefkopoulos, S. & Trompouki, E. Stress and non-stress roles of inflammatory signals during HSC emergence and maintenance. *Front. Immunol.* **7**, 487 (2016).
 20. Pietras, E. M. Inflammation: a key regulator of hematopoietic stem cell fate in health and disease. *Blood* **130**, 1693–1698 (2017).
 21. Harris, P. A. et al. Discovery of a first-in-class receptor interacting protein 1 (RIP1) kinase specific clinical candidate (GSK2982772) for the treatment of inflammatory diseases. *J. Med. Chem.* **60**, 1247–1261 (2017).
 22. Roderick, J. E. et al. Hematopoietic RIPK1 deficiency results in bone marrow failure caused by apoptosis and RIPK3-mediated necroptosis. *Proc. Natl Acad. Sci. USA* **111**, 14436–14441 (2014).
 23. Rickard, J. A. et al. RIPK1 regulates RIPK3-MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell* **157**, 1175–1188 (2014).
 24. Dillon, C. P. et al. Survival function of the FADD-CASPASE-8-cFLIP_L complex. *Cell Reports* **1**, 401–407 (2012).
- Acknowledgements** V. Dixit and K. Newton provided *Ripk3*^{−/−} mice, S. Hedrick and R. Hakem provided *Casp8*^{fl/fl} mice; H. Anderton and U. Nachbur helped with *Sharpin*^{cdpm} analysis; P. Levy and staff and L. Lawrence provided technical and histology services. A. Leister and J. Marinis advised on GSK'457A. T. Marafioti and A. Akarsa helped with phosphorylated MLKL detection. A. Annibaldi provided scientific advice and helpful discussions. B. J. Ferguson provided reagents and advice. The flow cytometry, microscopy and imaging core facilities are supported by Cancer Research UK through the CRUK-UCL Centre (515818) and a Cancer Immunotherapy Accelerator award (CITA, 525877). This work was funded by a Wellcome Trust Senior Investigator Award (096831/Z/11/Z), an ERC Advanced grant (294880) and a Cancer Research UK programme grant (A17341) awarded to H.W., NHMRC grants awarded to P.B., A. St., J.S. and H.W. (602516, 1037321, 1043414, 1080321, 1105209, 461221, 1042629, 1057905), the Leukemia and Lymphoma Society (Specialised Center of Research grant 7015) and a postdoctoral fellowship awarded to N.P. by the Swiss National Science Foundation (P300P3_158509).
- Reviewer information** Nature thanks F. Chan and the other anonymous reviewer(s) for their contribution to the peer review of this work.
- Author contributions** H.W. conceived the project. N.P. and M.D. performed most experiments. N.P., M.D. and H.W. designed the research and co-wrote the manuscript. A.M., C.B. and T.E. conceived and contributed to the haematopoietic analyses. H.D. and P.D. contributed to in vitro experiments in Fig. 2 and Extended Data Fig. 3. S.K. generated *Mkl*^{−/−} mice, L.T., E.R. contributed to in vivo experiments, T.H. performed cytokine arrays and E.L. and Y.S. contributed with biochemistry data. P.B., T.L.H. and H.W. designed the *Hoil-1*-floxed allele and P.B. generated it. A.F. and W.K. generated and analysed *Ripk1*^{−/−}*Ripk3*^{−/−}*Casp8*^{−/−}*Hoil-1*^{−/−} mice. H.D. and A.Sa. performed genotyping. A.B. and J.B. provided GSK'547A and *Ripk1*^{K45A} mice. J.R., S.A.-D., A.St. and J.S. performed the *Sharpin*^{cdpm} studies. C.H. and M.T.A. performed pathological and microfocus computed tomography analyses. T.E., P.B., A.St., J.S. and E.R. provided scientific insight.
- Competing interests** J.B. and A.L. are GSK employees.
- Additional information**
Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0064-8>.
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0064-8>.
Reprints and permissions information is available at <http://www.nature.com/reprints>.
Correspondence and requests for materials should be addressed to H.W.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mice. The *Hoil-1*-floxed (*Hoil-1^{fl/fl}*) mice were generated by a gene-targeting strategy in ES cells in which the targeting cassette was composed of a hygromycin-resistance cassette flanked by *Frt* sites and exons 1 and 2 of the *Hoil-1* gene flanked by *loxP* sites. Southern blots of C57BL/6 ES cell clones containing the homologous recombination were analysed for the specificity of the recombination and the absence of any unwanted integration. Two ES cell clones were used to generate mutant animals on the C57BL/6 genetic background, corresponding to the two independent *Hoil-1^{-/-}* strains (*Hoil-1^{-/-}* and C20*Hoil-1^{-/-}*). The hygromycin cassette was removed by crossing these mice with C57BL/6 mice expressing the FlpE recombinase and this was followed by a cross with C57BL/6 mice to remove the flpe transgene. *Hoip^{-/-}* and *Hoil-1^{-/-}* mice were generated by crossing *Hoip^{fl/fl}* mice, previously described¹⁷, and *Hoil-1^{fl/fl}* mice (described here) with transgenic mice expressing the *loxP*-deleter Cre recombinase (purchased from JAX: 6054, B6.C-Tg(CMV-Cre)1 Cgn/J). Transgenic mice expressing the Cre recombinase under the control of the *Tie2* (also known as *Tek*) promoter (*Tie2-Cre*) (B6.Cg-Tg(Tek-cre)1Ywa/J)²⁵ were used to delete floxed genes specifically in endothelial cells. C57BL/6 *Mkl1^{-/-}* mice crossed to *Sharpin^{cpdm}* mice were previously described²⁶. For all other crosses *Mkl1^{-/-}* mice were generated using transcription activator-like effector nuclease (TALEN). In brief, TALENs targeting exon 1 of the *Mkl1* gene were cloned via Golden-gate assembly. The RVD sequence of TAL1 against TACCGTTTCAGATGTCA was NIHDHNNNGNGNGHDNINNNINGNNGHNDNI, and TAL2 against TCGATCTTCCTGCTGCC was HDNNNNGHNDNGHNDHNDNGNNDHND. Capped RNA was produced in vitro using mMESSAGE mMACHINE T7 Transcription Kit (Ambion) and poly A tail was added using Poly(A) Tailing Kit (Ambion). Purified transcripts were mixed and adjusted to 25 ng μ l⁻¹. C57BL/6 fertilized eggs were injected into both the cytoplasm and the pro-nucleus. Embryos were transferred into C57BL/6 pseudo-pregnant females. Pups were genotyped by sequencing using genomic DNA obtained from ear punches. One female carrying a 19-base pair (bp) homozygous deletion causing a premature stop codon was selected for further breeding. *Mkl1^{-/-}* mice were backcrossed to C57BL/6 mice for two generations. *Sharpin^{cpdm}* (C57BL/Ka) and *Tnfr1^{-/-}* (2818, B6.129-Tnfrsf1atm1Mak/J) mice were purchased from JAX. *Tnf^{-/-}* mice (C57BL/6;129S6) were provided by W. Kaiser. *Ripk3^{-/-}*²⁷, *Casp8^{-/-}*²⁸, *Ripk1^{K54A}* and *Ripk1^{-/-}*²⁹ mice have been reported previously. Timed matings were performed as previously described⁸. All mice were genotyped by PCR, fed ad libitum. All animal experiments were conducted under an appropriate UK project license in accordance with the regulations of UK home office for animal welfare according to ASPA (animal (scientific procedure) Act 1986). The relevant Animal Ethics Committee approved all experiments involving *Sharpin^{cpdm}* and the *Ripk1^{-/-}* crosses which were maintained under appropriate licenses and subject to ethical review at The Walter and Eliza Hall Institute (Melbourne, Australia) and UT Health Sciences Center San Antonio (TX, USA), respectively.

Histological analysis, TUNEL and immunofluorescence staining. Embryos or organs from adult mice were collected and fixed in 10% buffered formalin and paraffin embedded. Sections of 4- μ m were stained with haematoxylin and eosin following standard procedures. Necropsy of adult mice or six sagittal serial sections of two different planes of the embryo were used for blinded pathological analysis. For TUNEL staining, sections were treated according to the manufacturer's instructions (DeadEnd Fluorometric TUNEL System, Promega, G3250). For whole-mount TUNEL staining and immunofluorescence staining, samples were processed using the ApoTag plus Peroxidase In situ Apoptosis Detection Kit (Millipore, S7101) according to the manufacturer's instructions and as previously described⁸. Quantification was performed by an experimenter blinded to the genotype of the mice by using ImageJ Software on monochrome images of the whole yolk sac by measuring the area of positive staining. Alternatively, TUNEL-positive cells were counted on five different fields (10 \times magnification). Yolk sacs were stained with antibodies against PECAM-1 (BD Biosciences, 5533370 clone MEC13.3) and cleaved caspase 3 (Cell Signaling, 9664), followed by staining with secondary antibodies, Alexa Fluor 594 goat anti-rat IgG and Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen, A-11007 and A-11034, respectively), and analysed by fluorescent microscopy. Quantification was performed by an experimenter blinded to the genotype of the mice on ten different fields (10 \times magnification) per yolk sac. **Microfocus computed tomography scan.** Embryos were fixed in 4% paraformaldehyde and potassium triiodide (Lugol's iodine/I₂KI, to impart tissue contrast), with a total iodine content of 63.25 mg ml⁻¹ (iodine mass of 2.49×10^{-4} mol ml⁻¹), in a 1:1 ratio for 8 h before imaging. Before scanning, the embryos were washed, wrapped in Parafilm M (Bemis) and secured in 3% (w/v) Agar (Sigma-Aldrich), within a low-density plastic cylinder to ensure mechanical stability during scan acquisition. Images were acquired using an XT H 225 ST microfocus-computed tomography scanner with a multimetal target (Nikon Metrology). Scans were reconstructed using modified Feldkamp filtered back projection algorithms with proprietary software (CTPro3D; Nikon Metrology) and post-processed using VG Studio

MAX (Volume Graphics GmbH). Soft tissues were analysed by Phong shading of direct volume renderings and plain projections and the vascular system by maximum intensity projections.

Cells. MEFs were isolated from E12.5–E13.5 embryos in accordance with standard procedures and these cells were maintained in DMEM medium supplemented with 10% fetal bovine serum (Sigma). Transformation was performed by lentiviral infection with the SV40 large T antigen. For reconstitution experiments, the coding sequence of mouse HOIP, SHARPIN or HOIL-1 wild-type (WT), the UBL domain of HOIL-1 only (HOIL-1-UBL; amino acids 1–139), HOIL-1- Δ RBR (amino acids 1–252), HOIL-1- Δ UBL (amino acids 140–508), HOIL-1 with inactivating mutations T201A/R208A in the NZF domain (HOIL-1-NZFmut) or HOIL-1 with a point mutation in the catalytic cysteine of the RBR domain (HOIL-1-C458A) was inserted in MSCV vector followed by the internal ribosome entry site (IRES)-GFP sequence. These vectors were retrovirally transduced into MEFs and GFP-positive cells were sorted in a MoFlo cytometer (Beckman Coulter).

Immunoprecipitation. For isolation of the TNFR1-SC, transformed MEFs were stimulated with 3 \times Flag-2 \times Strep-TNF at 0.5 μ g ml⁻¹ for 15 min, and controls were left untreated. Cells were subsequently solubilised in lysis buffer (30 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM EDTA, 2 mM KCl, 10% glycerol, 1% Triton X-100, EDTA-free proteinase inhibitor cocktail (Roche, 5056489001) and 1 \times phosphatase-inhibitor cocktail 2 (Sigma, P5726-1ML)) at 4°C for 30 min. The lysates were cleared by centrifugation, and 3 \times Flag-2 \times Strep-TNF (0.5 μ g ml⁻¹ per sample) was added to the untreated samples. Subsequently, the lysates were subjected to anti-Flag immunoprecipitation using M2 antibody coupled sepharose beads (Sigma, A2220-5ML) for 16 h. For immunoprecipitation of FADD, transformed MEFs were treated with 20 μ M zVAD-fmk (Abcam, ab120487) in the presence or absence of 100 ng ml⁻¹ 6 \times His-TNF for 3 h. Cells were lysed as described above and FADD was immunoprecipitated using anti-FADD antibody (Santa Cruz, sc-5559) and protein G Sepharose Beads (GE healthcare, 17-0618-01) at 4°C for 4 h. For SHARPIN immunoprecipitation, anti-SHARPIN antibody (ProteinTech, 14626-1-AP) was used. For all immunoprecipitations, the beads were washed three times with lysis buffer. Proteins were eluted in 50 μ l of LDS buffer (NuPAGE, Invitrogen) containing 50 mM dithiothreitol (DTT). Samples were analysed by western blotting.

Western blot analysis and antibodies. Whole embryos were snap-frozen and homogenized in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.5% sodium deoxycholate, 1% NP-40 and 1 \times EDTA-free proteinase inhibitor cocktail (Roche, 5056489001) or RIPA buffer with 6 M urea for the experiment in Extended data Fig. 7h. Alternatively, cells were washed twice with ice-cold PBS before lysis in lysis buffer. Protein concentration of lysates was determined using BCA protein assay (Thermo Scientific). Lysates were subsequently denatured in reducing sample buffer at 95°C for 10 min before separation by SDS-PAGE (NuPAGE) and subsequent analysis by western blotting using antibodies against HOIL-1³⁰, HOIP (custom-made, Thermo Fisher Scientific), SHARPIN (ProteinTech, 14626-1-AP), TNFR1 (Abcam, ab19139), actin (Sigma, A1978), pIKK α (Cell Signaling, 9246), IKK β (Cell Signaling, 9242), phospho p65 (Cell Signaling, 3033) cleaved caspase-8 (Cell Signaling, 9429), linear ubiquitin (Merck Millipore, MABS199), RIPK1 (BD, 610459), RIPK3 (Enzo, ADI-905-242-100), FADD (Assay Design, AAM-121), MLKL (Millipore, MABC604), phospho-MLKL (Abcam, ab196436) and tubulin (Sigma, T9026).

Cell death analysis by propidium iodide staining. Cells were seeded to 80% confluence and were then incubated with 100 ng ml⁻¹ His-tagged TNF, 1 μ g ml⁻¹ CD95L-Fc, 1 μ g ml⁻¹ isoleucine zipper tagged murine TRAIL (iz-mTRAIL), 100 μ g ml⁻¹ poly(I:C) HMW (InvivoGen, thrl-pic), 20 ng ml⁻¹ IFN- γ (Peprotech, 315-05) or 100 ng ml⁻¹ LT- α (Thermo Fisher Scientific, 10270-HNAE) for 24 h, unless otherwise indicated. When indicated the following inhibitors were used: 20 μ M zVAD-FMK (Abcam, ab120487), 10 μ M necrostatin-1 (Biovision, 2263-5). Supernatants and adherent cells were collected and resuspended in PBS containing 5 μ g ml⁻¹ propidium iodide. Propidium iodide-positive cells were enumerated by FACS (BD Accuri).

RNA-sequencing analysis. E13.5 embryos were snap frozen and RNA was prepared using the RNeasy minikit (Qiagen, 74104) according to the manufacturer's instruction. To generate the library, samples were processed using the KAPA mRNA HyperPrep Kit (KK8580) according to the manufacturer's instructions. In brief, mRNA was isolated from total RNA using Oligo dT beads to pull down poly-adenylated transcripts. The purified mRNA was fragmented using chemical fragmentation (heat and divalent metal cation) and primed with random hexamers. Strand-specific first strand cDNA was generated using reverse transcriptase in the presence of actinomycin D. The second cDNA strand was synthesized using dUTP in place of dTTP, to mark the second strand. The resultant cDNA was then 'A-tailed' at the 3' end to prevent self-ligation and adaptor dimerization. Truncated adaptors, containing a T overhang were ligated to the A-tailed cDNA. Successfully ligated cDNA molecules were then enriched with limited cycle PCR. Libraries to be multiplexed in the same run were pooled in equimolar quantities, calculated from Qubit

and Bioanalyser fragment analysis. Samples were sequenced on the NextSeq 500 instrument (Illumina) using a 43-bp paired-end run. Run data were de-multiplexed and converted to fastq files using the Illumina bcl2fastq Conversion Software v2.18 on BaseSpace. Fastq files were then aligned to a reference genome using STAR on the BaseSpace RNA-Seq alignment app v1.1.0. Reads per transcript were counted using HTSeq and differential expression was estimated using the BioConductor package DESeq2 (BaseSpace app v1.0.0). Next, four groups of differentially regulated genes were analysed: low and high abundance *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} versus *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} embryos and low and high abundance in *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} versus *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos. To identify genes that were specifically altered in the absence of HOIL-1, the Venny 2.1 software was used to exclude genes that were differentially expressed between *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} embryos from those between *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos. Genes that were already differentially expressed between the corresponding HOIL-1-expressing controls (that is, *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} embryos) were excluded from the differentially expressed genes between *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos. The resulting list of genes (33/85) was entered in the STRING software (<https://string-db.org>) to assess for functional enrichment in biological networks. Gene Ontology (GO) terms with false discovery rate below 1% are shown.

Flow cytometry analysis, colony-forming unit assay and macrophage culture.

For phenotypic analysis, single-cell suspensions from mechanically dissociated E13.5 fetal livers or a pool of aortas (AGM region) from three embryos were stained for 30 min on ice with various antibody cocktails. The antibodies against the surface markers examined were: CD16/32, clone 93 and 2.4G2 (eBioscience, 45-0161-82 and BD553141), CD135, clone A2F10.1 (BD, 553842), Ly-6A/E, clone D7 (Sca-1) (BD, 558162), CD117 (c-Kit), clone 2B8 (BD, 560185), CD34, clone RAM34 (BD, 562608), mouse lineage cocktail, clones 17A2/RB6-8C5/RA3-6B2/Ter-119/M1/70 (Biolegend, 133313 and BD, 561301), CD16/32, clone 2.4G2 (BioXcell, CUS-HB-197), CD11b, clone M1/70 (Biolegend, 101228 and eBioscience, 15-0112-81), CD11c, clone HL3 (BD, 561241), F4/80, clone BM8 (Biolegend, 123110), GR-1, clone RB6-8C5 (Biolegend, 108416 and 108410), CD45, clone 30-F11 (Biolegend, 103128 and Biolegend, 103112), CD3ε, clone, 145-2C11 (Biolegend, 100310), B220, clone RA3-6B2, (Biolegend, 103210), CD71, clone RI7217 (Biolegend, 113807), TER-119, clone TER-119 (Biolegend, 116234) and fixable viability dye (eBioscience, 65-0864-18 and 65-0867-14). The myeloid progenitors were identified in the LK population as CD34⁺CD16/32⁻ (CMP), CD34⁺CD16/32⁺ (GMP); CD34⁺CD16/32⁻ (MEP). Fluorescence minus one (FMO) was used as a gating control. For quantification of absolute number of cells, a defined number of flow cytometric reference beads (Invitrogen) were mixed with the samples before acquisition. Samples were processed either using LSR Fortessa (BD Biosciences) or sorted in a FACS Aria FUSION cell sorter (BD Biosciences). Data were analysed with FlowJo 7.6.1 software (Treestar). Cytospin preparations of 10,000 cells per slide of E13.5 fetal liver homogenates were stained by May-Grünwald Giemsa staining and enucleated erythrocytes were quantified blindly as number of cells per HPF using ImageJ Software. For growth of primitive erythroid progenitor cells or all haematopoietic stem cells, 5,000 sorted Lin⁻c-KIT⁺ E13.5 liver cells were cultured in MethoCult SF containing cytokines, including EPO (Stem Cell, M3436) or Mouse Methylocellulose Complete Media (R&D, HSC007), respectively. Colonies were enumerated after 14 days of incubation. For preparation of fetal liver-derived macrophages, equal amounts of E13.5 single cell suspensions were cultured and differentiated for 5 days in DMEM supplemented with 10% FCS plus

20% L929-conditioned medium (as a source of M-CSF) supplemented or not with the indicated inhibitors. Cells were imaged using EVOS Auto cell imaging system and viability was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega, G7572). Alternatively, cells were stained with Hoechst dye and enumerated using Citation cell imaging platform.

Cytokine analysis. Embryo homogenates prepared as described above ('Western blot and antibodies' section) were analysed with Proteome Profiler Arrays (Mouse Angiogenesis Array, ARY015, and Mouse Cytokine array Panel A, ARY006 both R&D). ELISA kits used were the CXCL4 (R&D, DY595), CXCL11 (Abcam, ab204519), CXCL10 (R&D, DY466-05), IFN lambda 2/3 (Pbl assay science, 62830-1), IL-1β (ThermoFisher, BMS6002) and IFN-β ELISA (ThermoFisher, 424001).

Epidermal thickness quantification. Per mouse, 1–2 pieces of skin were taken and epidermal thickness was measured by microscopy using a 20× magnification. Quantification was performed by an experimenter blinded to the genotype of the mice by using the CellSens software with at least 20 measurements per mouse.

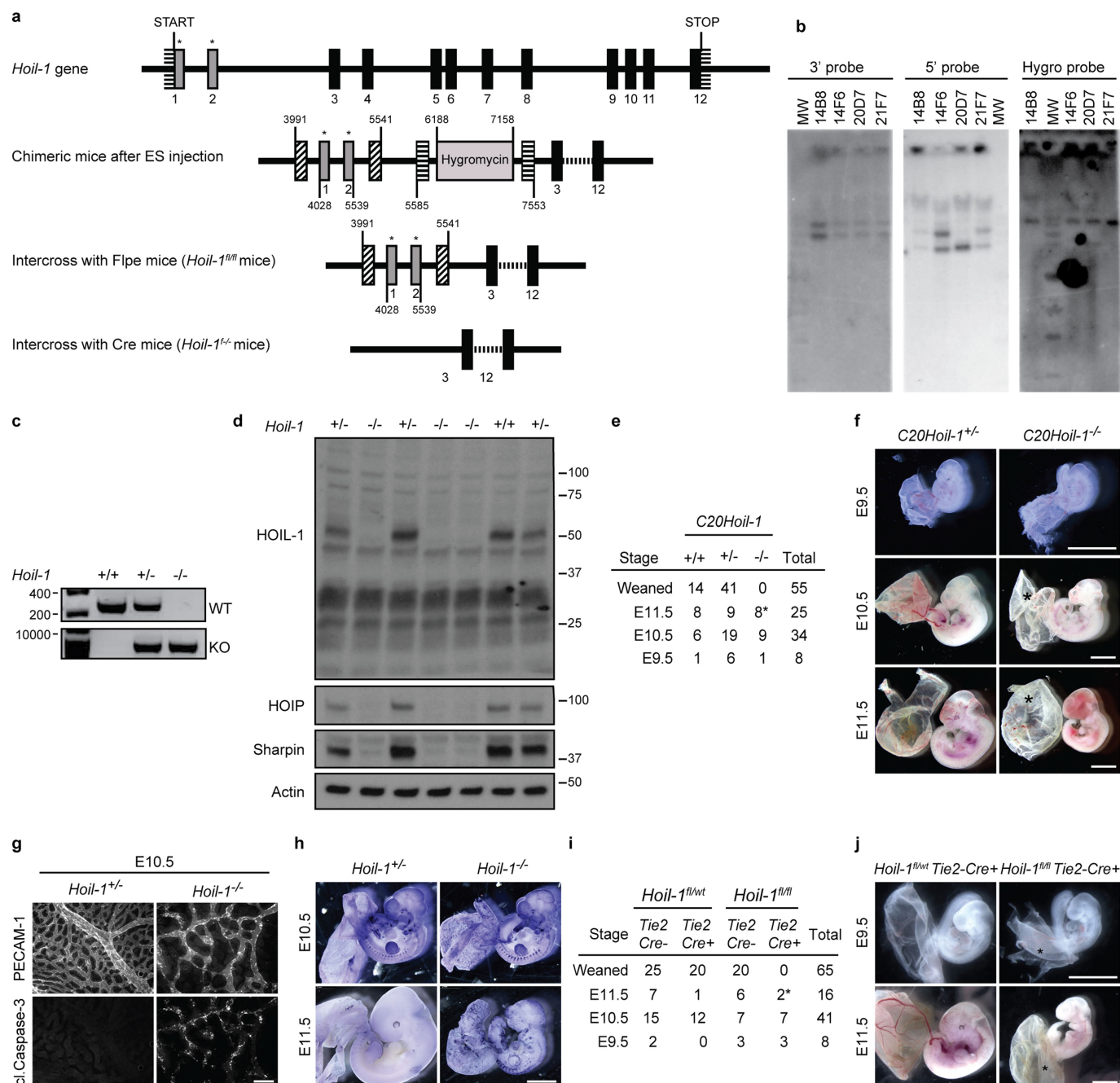
Pharmacological inhibition of RIPK1 kinase activity. Mice were fed with rodent chow containing 100 mg kg⁻¹ of the RIPK1 kinase inhibitor GSK3540547A (GSK'547A) (GlaxoSmithKline LLC) starting a week before mating and kept on this diet throughout pregnancy until caesarean section at the indicated time points.

Statistics and reproducibility. Group size was determined based on preliminary datasets. Statistical significance was determined using unpaired, two-tailed parametric Student's *t*-test. One- or two-way ANOVA with Tukey's multiple comparisons test was applied. 95% Confidence interval was considered for statistics and *P* < 0.05 was considered significant. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001. Multiplicity-adjusted *P* values are reported for multiple comparisons. All statistical analyses were performed using Graphpad Prism 6. Statistical transformations for RNA-seq were performed with DESeq2 and adjusted *P* values used the Benjamini–Hochberg test. All in vitro experiments were performed at least twice with similar results. Unless indicated in figure legends in vivo experiments were performed with at least two embryos per genotype. At least three embryos were considered for statistical testing. The experiments were not randomized.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. RNA-sequencing analysis data are available from the Sequence Read Archive (SRA) database SRP134865 (BioProject accession PRJNA437851) and comparative datasets including genes differentially regulated genes between embryo homogenates with different mutations are displayed in Supplementary Table 1.

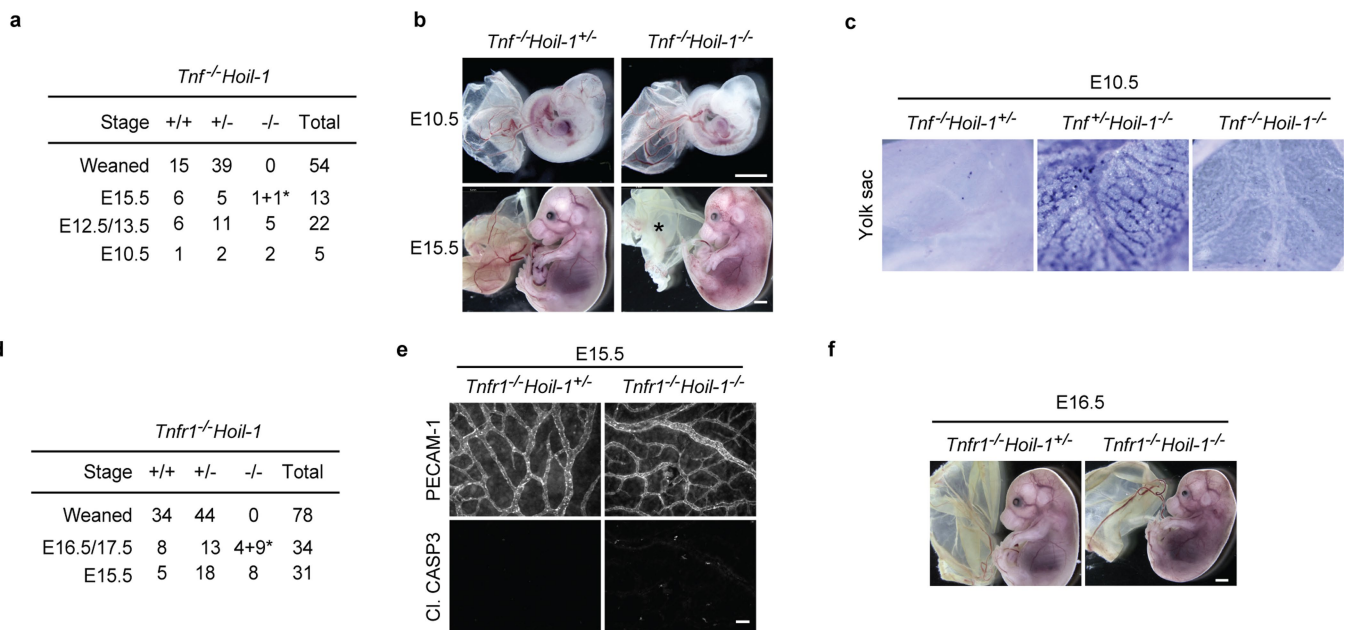
25. Gustafsson, E., Brakebusch, C., Hietanen, K. & Fassler, R. Tie-1-directed expression of Cre recombinase in endothelial cells of embryoid bodies and transgenic mice. *J. Cell Sci.* **114**, 671–676 (2001).
26. Murphy, J. M. et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* **39**, 443–453 (2013).
27. Newton, K., Sun, X. & Dixit, V. M. Kinase RIP3 is dispensable for normal NF-κBs, signaling by the B-cell and T-cell receptors, tumor necrosis factor receptor 1, and Toll-like receptors 2 and 4. *Mol. Cell. Biol.* **24**, 1464–1469 (2004).
28. Salmena, L. et al. Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. *Genes Dev.* **17**, 883–895 (2003).
29. Kelliher, M. A. et al. The death domain kinase RIP mediates the TNF-induced NF-κappaB signal. *Immunity* **8**, 297–303 (1998).
30. Haas, T. L. et al. Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction. *Mol. Cell* **36**, 831–844 (2009).



Extended Data Fig. 1 | HOIL-1-deficient mice die at mid-gestation.

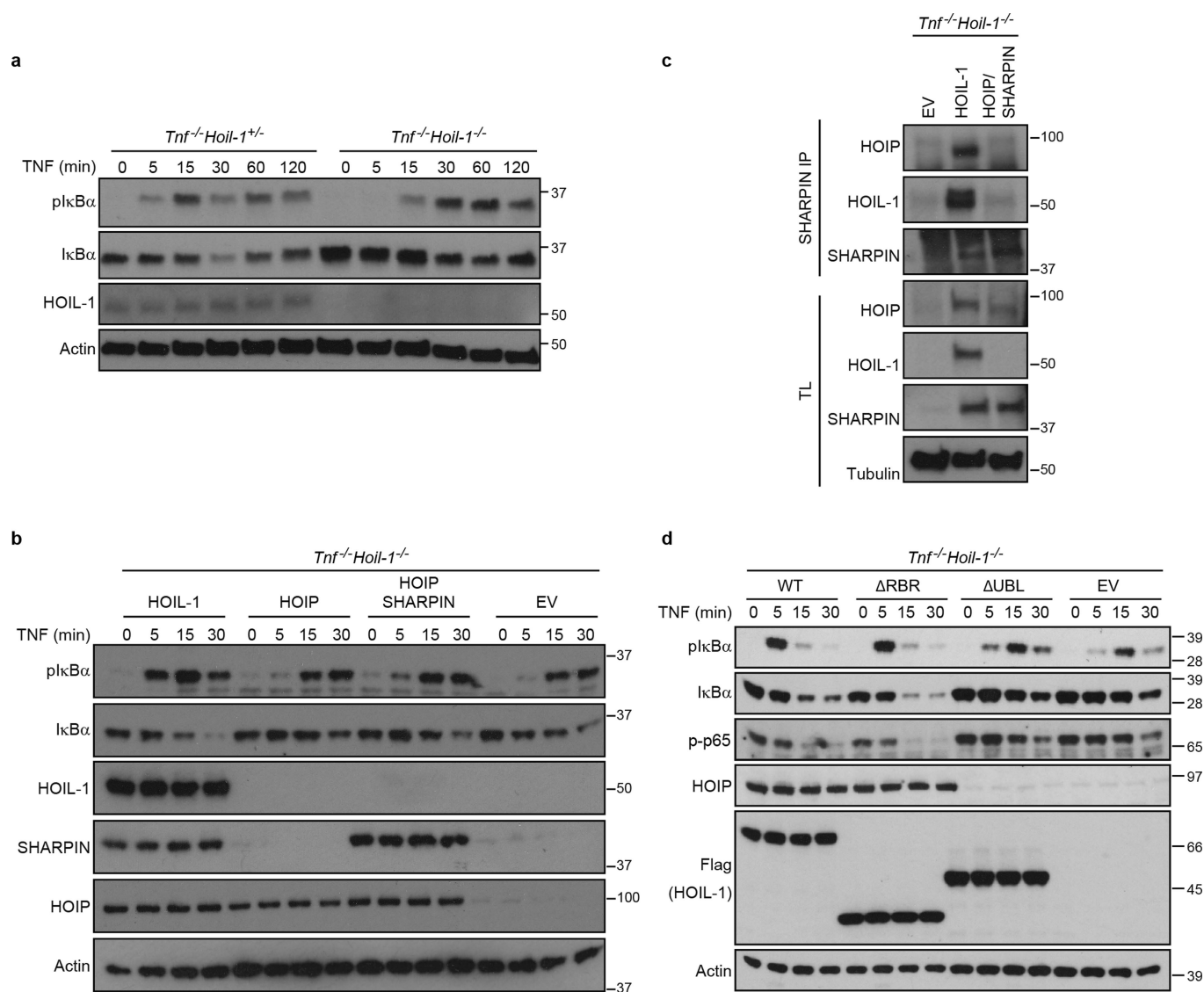
a, Schematic representation of the *Hoil-1*-knockout strategy. Solid boxes represent *Hoil-1* exons and grey boxes with a star indicate the targeted exons. Boxes with diagonal and horizontal strips represent *loxP* and *Frt* sites, respectively. **b**, Specificity of gene recombination was assessed by Southern blotting with 5' and 3' probes external to the construct in four clones (14B8, 14F6, 20D7 and 21F7). Digest of the DNA with *Apa*I, followed by hybridization with the 3' probe was expected to show a 5,700-bp band for the wild-type allele and a 7,700-bp band for the mutant allele. All four clones appeared to have the correct recombination on the 3' side. Digest of the DNA with *Sph*I and hybridization with the 5' probe was expected to show a 4,500-bp wild-type band and a 6,200-bp band for the mutated allele. Clones 14B8, 14F6 and 21F7 appeared to be correctly recombined on the 5' side. Finally, cutting the DNA with *Apa*I and hybridizing with a hygromycin probe showed a single band in all clones, indicative of a single integration of the construct in all four ES clones. Clones 14B8 and 14F6 were selected for generation of the two

Hoil-1^{-/-} strains. **c**, PCR analysis of *Hoil-1* wild-type, heterozygous and knockout mice. **d**, Protein levels of HOIL-1, HOIP and SHARPIN in whole embryo lysates ($n = 3$ for *Hoil-1*^{+/+} and *Hoil-1*^{-/-} embryos and $n = 1$ for *Hoil-1*^{+/-} embryos). For gel source data (c, d), see Supplementary Fig. 1. **e**, Quantification of genotypes of animals obtained from intercrossing C20*Hoil-1*^{+/+} mice. Asterisk indicates dead embryo. **f**, Representative images of C20*Hoil-1*^{+/+} and C20*Hoil-1*^{-/-} embryos from E9.5 to E11.5 as quantified in e. Scale bars, 2 mm. **g**, Single staining showing vascularization (PECAM-1, top) and apoptosis (cleaved CASP3, bottom) of yolk sacs. Merged image is shown in Fig. 1c. **h**, Whole-mount TUNEL staining of embryos at the indicated stages (embryo per genotype $n = 2$ at E10.5, $n = 8$ for *Hoil-1*^{+/+} and $n = 5$ for *Hoil-1*^{-/-} at E11.5). Scale bar, 2 mm. **i**, Quantification of genotypes of animals obtained from intercrossing *Hoil-1*^{fl/wt}*Tie2-cre*⁺ with *Hoil-1*^{fl/fl}*Tie2-cre*⁻ mice. Asterisk indicates dead embryo. **j**, Representative images of embryos with conditional deletion of *Hoil-1* in *Tie2-Cre*-expressing cells as quantified in i. Scale bar, 2 mm. Asterisks denote poorly vascularized yolk sac.



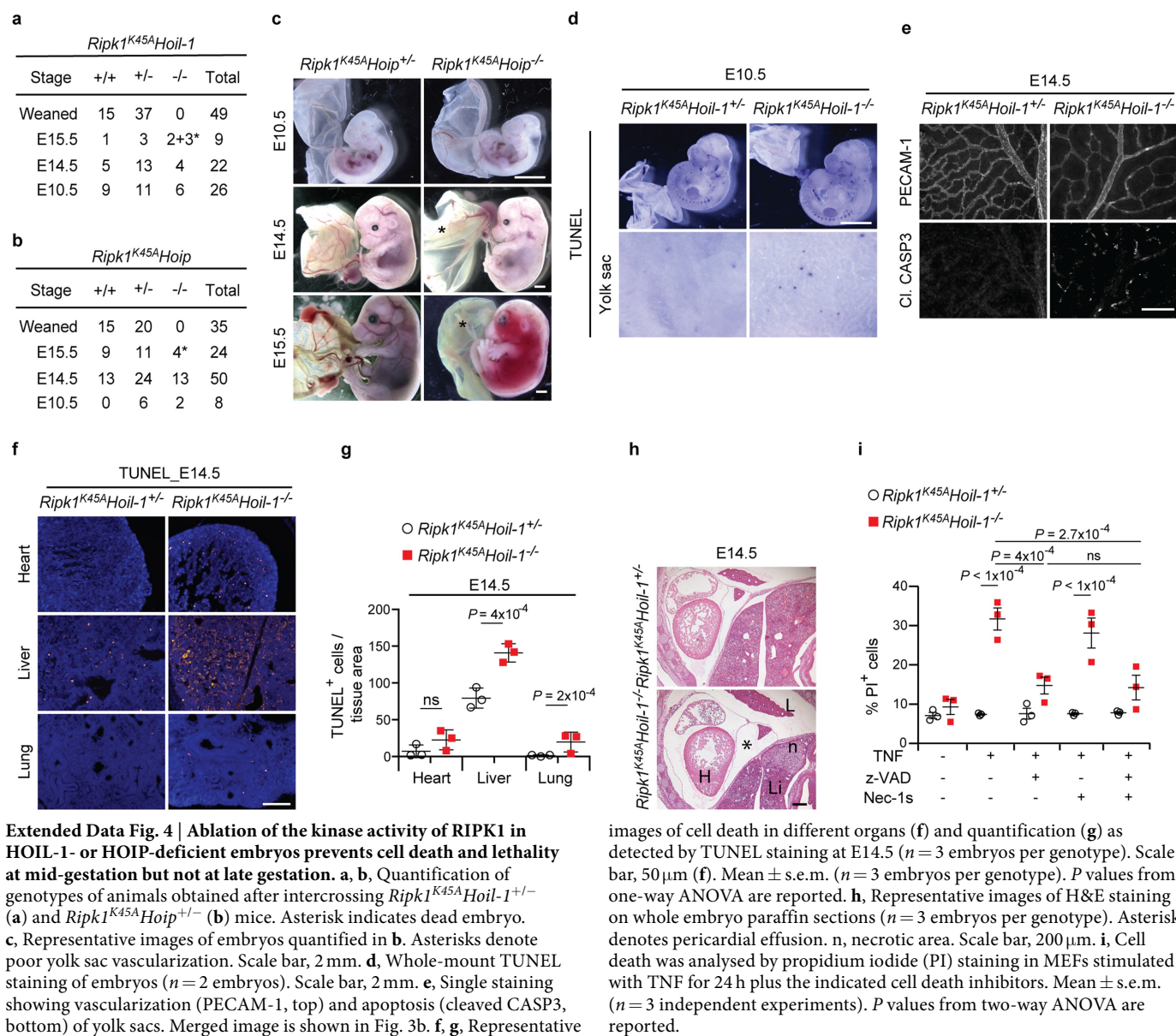
Extended Data Fig. 2 | TNFR1 signalling drives cell death and lethality of HOIL-1-deficient mice at mid-gestation. a, d, Quantification of genotypes of animals obtained from intercrosses of *Tnf^{-/-}Hoil-1^{+/-}* (a) and *Tnfr1^{-/-}Hoil-1^{+/-}* (d) mice. Asterisk denotes dead embryo. b, Representative images of embryos quantified in a at E10.5 and E15.5. Asterisk denotes poor yolk sac vascularization. c, Cell death as detected

by whole-mount TUNEL staining in yolk sacs at E10.5 ($n = 3$ embryos per genotype). e, Single staining showing vascularization (PECAM-1, top) and apoptosis (cleaved CASP3, bottom) of yolk sacs. Merged image is shown in Fig. 1g. Scale bar, 50 μ m. f, Representative images of embryos at E16.5 ($n = 2$ for *Tnfr1^{-/-}Hoil-1^{+/-}* and $n = 4$ for *Tnfr1^{-/-}Hoil-1^{-/-}*).



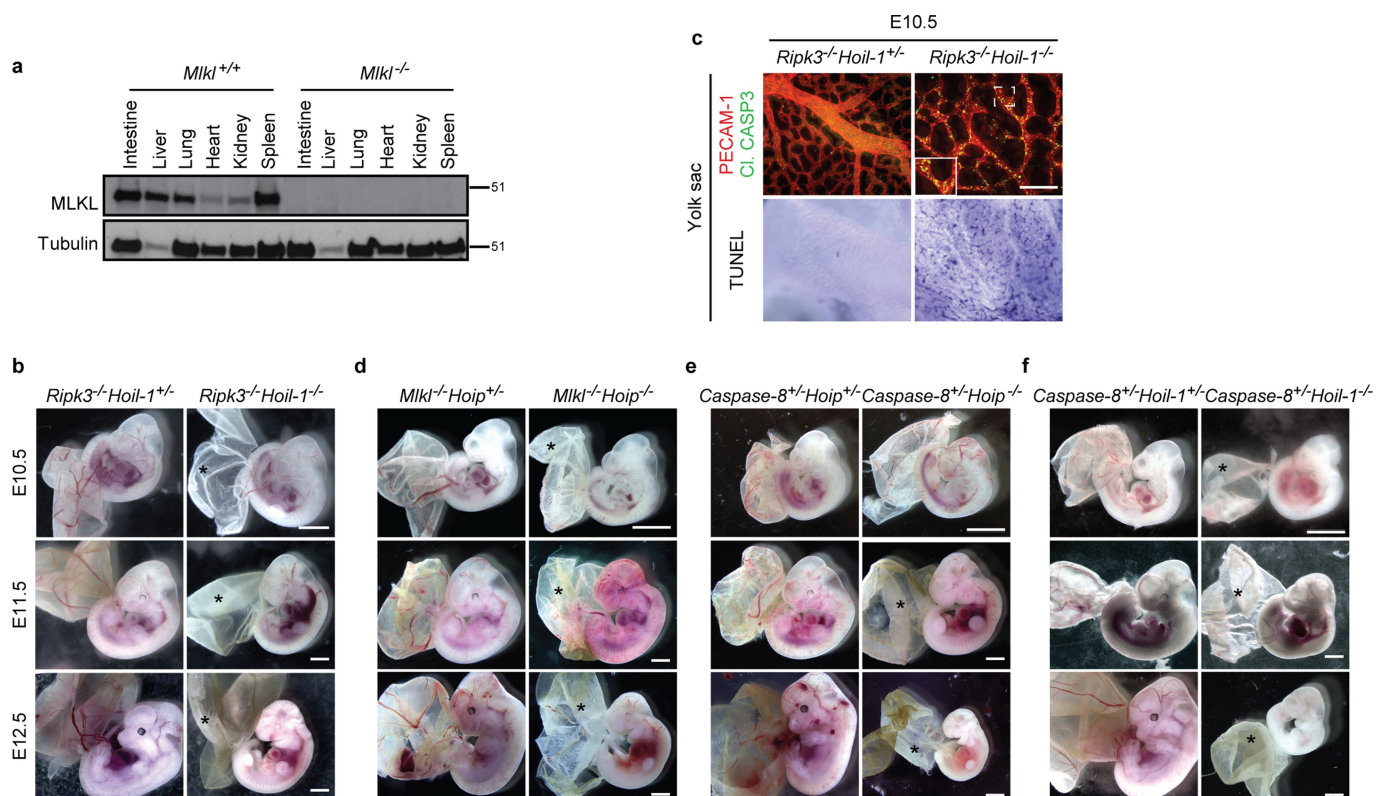
Extended Data Fig. 3 | HOIL-1 is required for optimal TNF-induced NF- κ B activation independently of its RBR domain. **a, **b**, **d**, Western blot analysis of the indicated proteins in whole-cell lysates from MEFs of the indicated genotypes after they had been stimulated with TNF (or left untreated) for the indicated time points (**a**), after overexpression of the different LUBAC components (**b**; HOIL-1, HOIP, SHARPIN) or**

after expression of the indicated mutant forms of HOIL-1 (d**) ($n = 2$ independent experiments). **c**, SHARPIN immunoprecipitation was performed in $Tnf^{-/-} Hoil-1^{-/-}$ MEFs reconstituted with HOIL-1 or a combination of HOIP and SHARPIN and analysed by western blotting ($n = 2$ independent experiments). For gel source data, see Supplementary Fig. 1.**



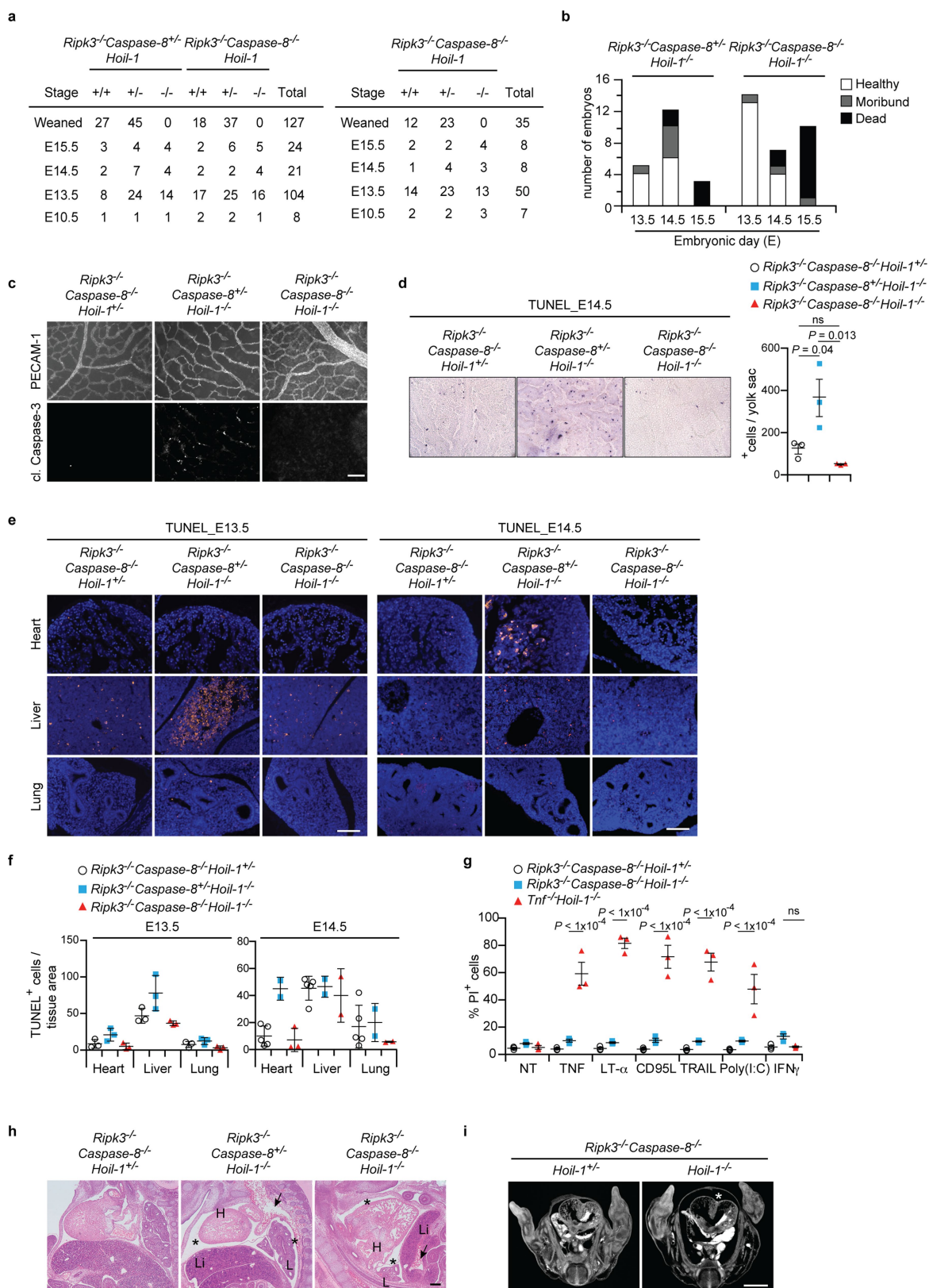
Extended Data Fig. 4 | Ablation of the kinase activity of RIPK1 in HOIL-1- or HOIP-deficient embryos prevents cell death and lethality at mid-gestation but not at late gestation. **a, b**, Quantification of genotypes of animals obtained after intercrossing *Ripk1^{K45A}Hoil-1^{+/-}* (**a**) and *Ripk1^{K45A}Hoip^{+/-}* (**b**) mice. Asterisk indicates dead embryo. **c**, Representative images of embryos quantified in **b**. Asterisks denote poor yolk sac vascularization. Scale bar, 2 mm. **d**, Whole-mount TUNEL staining of embryos ($n = 2$ embryos). Scale bar, 2 mm. **e**, Single staining showing vascularization (PECAM-1, top) and apoptosis (cleaved CASP3, bottom) of yolk sacs. Merged image is shown in Fig. 3b. **f, g**, Representative

images of cell death in different organs (**f**) and quantification (**g**) as detected by TUNEL staining at E14.5 ($n = 3$ embryos per genotype). Scale bar, 50 μ m (**f**). Mean \pm s.e.m. ($n = 3$ embryos per genotype). P values from one-way ANOVA are reported. **h**, Representative images of H&E staining on whole embryo paraffin sections ($n = 3$ embryos per genotype). Asterisk denotes pericardial effusion. n, necrotic area. Scale bar, 200 μ m. **i**, Cell death was analysed by propidium iodide (PI) staining in MEFs stimulated with TNF for 24 h plus the indicated cell death inhibitors. Mean \pm s.e.m. ($n = 3$ independent experiments). P values from two-way ANOVA are reported.



Extended Data Fig. 5 | Individual deletion of mediators of apoptosis or necroptosis does not prevent cell death and lethality at mid-gestation of HOIL-1- or HOIP-deficient embryos. **a**, Western blot analysis of MLKL expression in the indicated organs derived from *Mkl1*^{-/-} mice ($n = 2$ mice per genotype), as control. For gel source data, see Supplementary Fig. 1. **b**, **d**–**f**, Representative images of embryos at different stages of gestation (E10.5: $n = 7$ for *Ripk3*^{-/-}*Hoil-1*^{+/+} and $n = 5$ for *Ripk3*^{-/-}*Hoil-1*^{-/-}; E11.5: $n = 5$ for *Ripk3*^{-/-}*Hoil-1*^{+/+} and $n = 2$ for *Ripk3*^{-/-}*Hoil-1*^{-/-}; E12.5: $n = 9$ for *Ripk3*^{-/-}*Hoil-1*^{+/+} and $n = 2$ for *Ripk3*^{-/-}*Hoil-1*^{-/-} (**b**), E10.5: $n = 16$ for *Mkl1*^{-/-}*Hoip*^{+/+} and $n = 6$ for *Mkl1*^{-/-}*Hoip*^{-/-}; E11.5: $n = 8$ for *Mkl1*^{-/-}*Hoip*^{+/+} and $n = 6$ for *Mkl1*^{-/-}*Hoip*^{-/-}; E12.5:

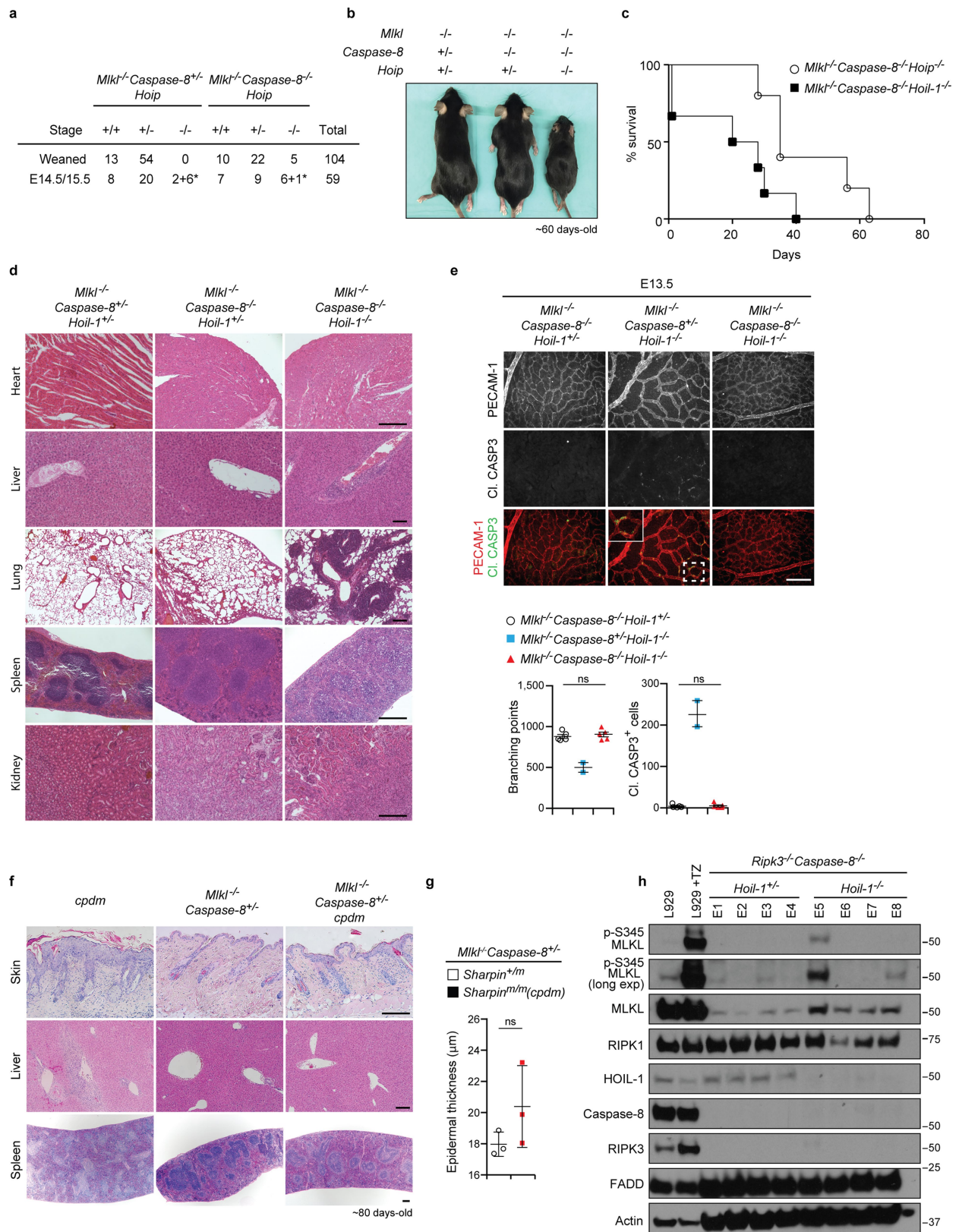
$n = 10$ for *Mkl1*^{-/-}*Hoip*^{+/+} and $n = 5$ for *Mkl1*^{-/-}*Hoip*^{-/-} (**d**), E10.5: $n = 5$ for *Casp8*^{+/+}*Hoip*^{+/+} and $n = 4$ for *Casp8*^{+/+}*Hoip*^{-/-}; E11.5: $n = 6$ for *Casp8*^{+/+}*Hoip*^{+/+} and $n = 3$ for *Casp8*^{+/+}*Hoip*^{-/-}; E12.5: $n = 3$ for *Casp8*^{+/+}*Hoip*^{+/+} and $n = 2$ for *Casp8*^{+/+}*Hoip*^{-/-} (**e**), E10.5: $n = 2$ for *Casp8*^{+/+}*Hoil-1*^{+/+} and $n = 4$ for *Casp8*^{+/+}*Hoil-1*^{-/-}; E11.5: $n = 2$ for *Casp8*^{+/+}*Hoil-1*^{+/+} and $n = 5$ for *Casp8*^{+/+}*Hoil-1*^{-/-}; E12.5: $n = 6$ for *Casp8*^{+/+}*Hoil-1*^{+/+} and $n = 3$ for *Casp8*^{+/+}*Hoil-1*^{-/-} (**f**). Asterisks denote poor yolk sac vascularization and cell death at E10.5 as detected by PECAM-1 (red) and cleaved CASP3 staining (green) (top) and whole-mount TUNEL staining (bottom) ($n = 4$ per genotype). Scale bar, 50 μm .



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Combined deletion of RIPK3 and caspase-8 prevents cell death but not embryonic lethality at late gestation that is caused by the loss of HOIL-1. **a**, Quantification of genotypes of animals obtained from inter-crosses of *Ripk3*^{-/-}*Casp8*^{+/-}*Hoil-1*^{+/-} with *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} mice (left) or *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} mice (right). **b**, Health status of *Ripk3*^{-/-}*Casp8*^{+/-}*Hoil-1*^{-/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos at different developmental stages. **c**, Single staining showing vascularization (PECAM-1, top) and apoptosis (cleaved CASP3, bottom) of yolk sacs. Merged image is shown in Fig. 3f. Scale bar, 50 μ m. **d**, Cell death as detected by whole-mount TUNEL staining in yolk sacs at E14.5 (left) and respective quantification (right). Mean \pm s.e.m. ($n = 3$ embryos per genotype). *P* values from one-way ANOVA are reported. **e, f**, Representative images (**e**) and quantification (**f**) of cell death in different organs as detected by TUNEL staining at E13.5 ($n = 3$ embryos per

genotype) and E14.5 ($n = 5$ for *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-}, $n = 2$ for *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} lung and liver and $n = 3$ *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} heart). Scale bars, 50 μ m. Data are mean \pm s.e.m. **g**, Cell death was analysed by propidium iodide (PI) staining in MEFs stimulated or not with the indicated ligands for 24 h. Data are mean \pm s.e.m. ($n = 3$ independent experiments). *P* values from two-way ANOVA are reported. **h**, Representative images of H&E staining on E13.5 whole embryo paraffin embedded sections ($n = 3$ for *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and $n = 2$ for *Ripk3*^{-/-}*Casp8*^{+/-}*Hoil-1*^{-/-}). Asterisks denote pericardial effusion. Arrows denote congested vessels. Scale bar, 200 μ m. **i**, Representative images of microfocus computed tomography scan images of whole E13.5 embryos ($n = 3$ embryos per genotype). Asterisks denote pericardial effusion.

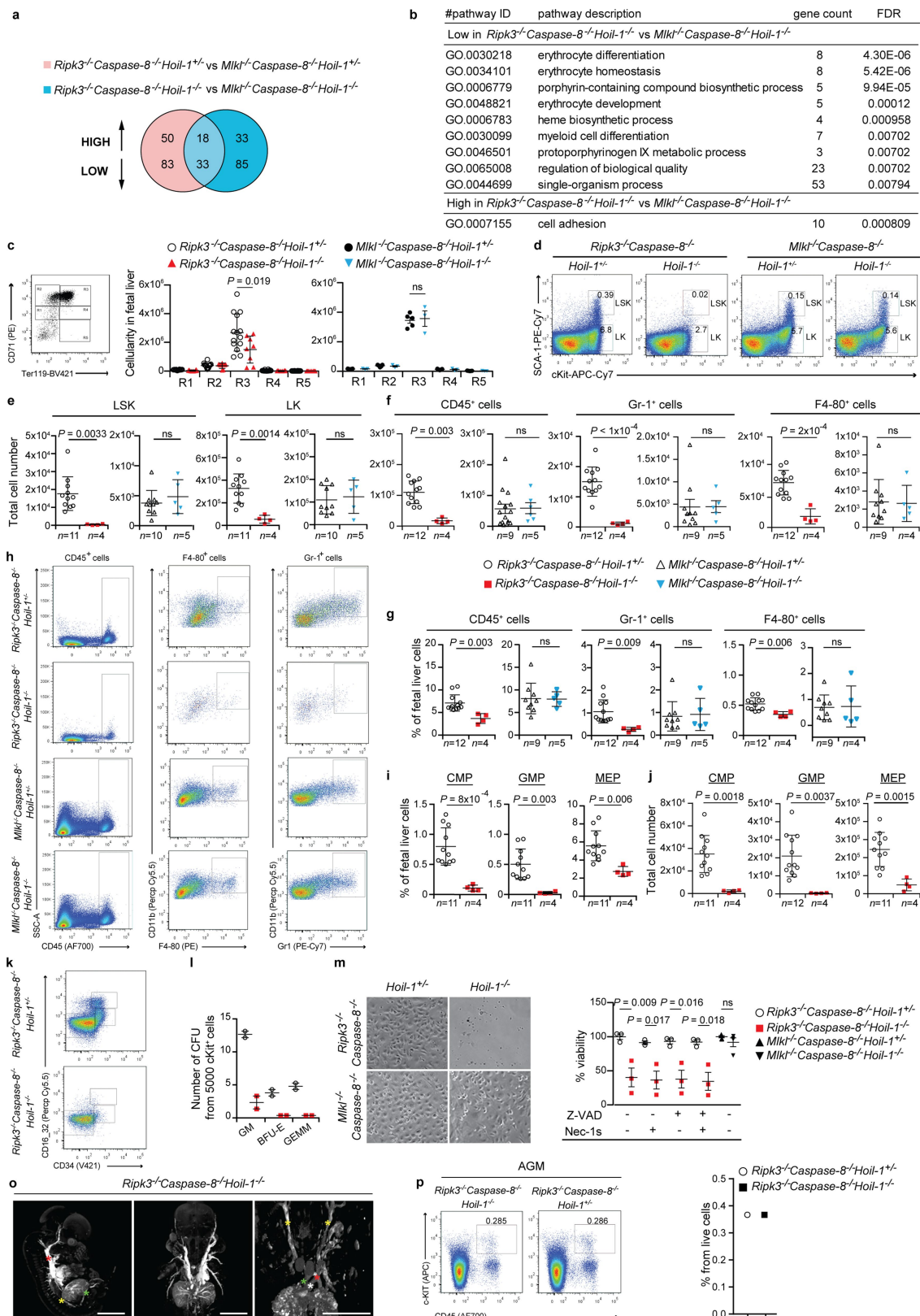


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Combined deletion of MLKL and caspase-8 promotes survival of LUBAC-deficient mice.

a. Quantification of genotypes of animals obtained from intercrosses of *Mkl^{-/-}Casp8^{+/-}Hoip^{+/-}* with *Mkl^{-/-}Casp8^{-/-}Hoip^{+/-}* mice. Asterisk denotes dead embryo. **b.** Representative images of adult mice as quantified in **a**. **c.** Kaplan–Meier plot of mouse survival ($n = 6$ for *Mkl^{-/-}Casp8^{-/-}Hoip^{-/-}* and $n = 9$ for *Mkl^{-/-}Casp8^{-/-}Hoil-1^{-/-}* mice). **d.** Representative images of H&E staining of the indicated organs ($n = 3$ mice per genotype). Scale bars, 200 μm . **e.** Representative images of yolk sac vascularization (PECAM-1, red) and apoptosis (cleaved CASP3, green) (top) at E13.5 and respective quantifications (bottom). Data are mean \pm s.e.m. ($n = 5$ for *Mkl^{-/-}Casp8^{-/-}Hoil-1^{+/-}* and

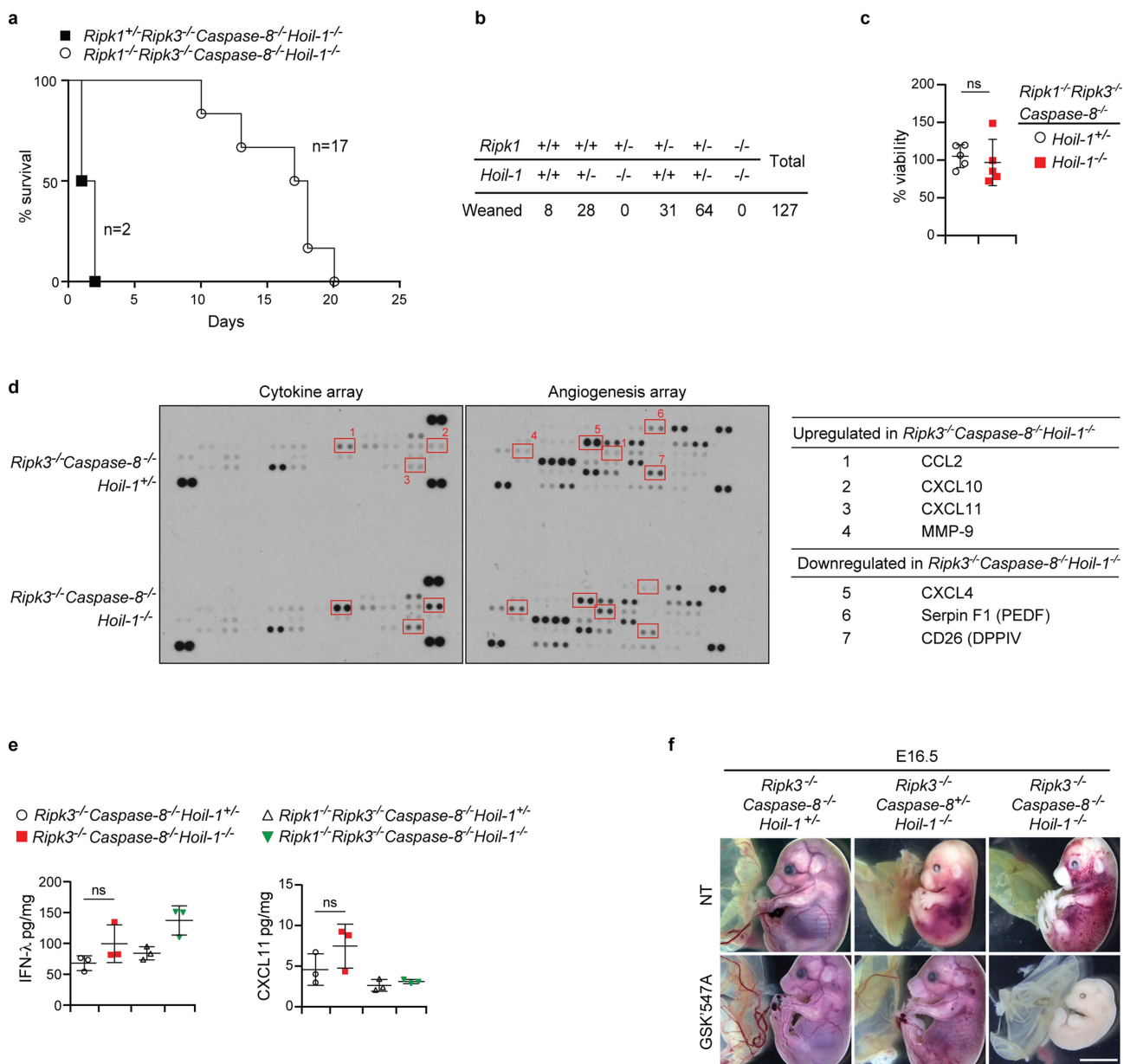
Mkl^{-/-}Casp8^{-/-}Hoil-1^{-/-} and $n = 2$ for *Mkl^{-/-}Casp8^{+/-}Hoil-1^{-/-}*). Statistical significance was determined with unpaired two-tailed *t*-tests comparing *Mkl^{-/-}Casp8^{-/-}Hoil-1^{+/-}* and *Mkl^{-/-}Casp8^{-/-}Hoil-1^{-/-}* embryos. **f.** Representative H&E staining images of the indicated organs ($n = 3$ embryos per genotype). Scale bars, 200 μm . **g.** Epidermal thickness quantification of mice of the indicated genotypes in **f**. Data are mean \pm s.e.m. ($n = 3$ mice per genotype). Statistical significance was determined with unpaired two-tailed *t*-tests. **h.** Western blot analysis of lysates from whole E13.5 embryos of the indicated genotypes and L929 cells treated with/without TNF plus zVAD-fmk (TZ) for 2 h as antibody validation ($n = 4$ embryos per genotype performed twice). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Combined deletion of RIPK3 and caspase-8 causes haematopoietic defects and RIPK1-dependent embryonic lethality in HOIL-1-deficient mice. **a**, Venn diagram depicting genes differentially expressed by RNA-seq analysis between E13.5 embryos of the indicated genotypes. **b**, Gene Ontology (GO) enrichment analysis of differentially expressed genes (85 low and 35 high in **a**). FDR, false discovery rate. **c**, Representative FACS profile of E13.5 fetal liver cells with different erythroblast populations gated according to their CD71 and TER119 expression levels (R1–R5) and quantification. R1 contains immature red blood cell progenitors, including primitive and later-stage erythroid progenitor cells (erythroid burst-forming unit (BFU-E) and colony-forming unit (CFU-E), respectively); R2 comprises mainly pro-erythroblasts and early basophilic erythroblasts; R3 contains both early and late basophilic erythroblasts; R4 is composed of chromatophilic and orthochromatophilic erythroblasts; and R5 consists of late orthochromatophilic erythroblasts and reticulocytes. Data are mean \pm s.e.m. ($n = 14$ *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-}, $n = 8$ *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-}, $n = 5$ for *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} and $n = 3$ for *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} fetal livers). *P* values from two-way ANOVA are reported. **d**, **h**, **k**, Representative FACS profile of E13.5 fetal liver cells for the indicated haematopoietic populations (sample size specified in **e–g**, **i**, **j**). **e**, **f**, **j**, Total cell number of the different haematopoietic cell subsets in fetal liver cell suspensions from E13.5 embryos of the indicated genotypes gated as in **d**, **h** and **k**, respectively. Total number of multipotent progenitors (LSK and LK cells) (**e**), mature CD45⁺ blood cells, including granulocytes (GR-1⁺) and macrophages (F4-80⁺) (**f**) and myeloid progenitors (common myeloid progenitor (CMP), granulocyte-monocyte progenitor (GMP) and megakaryocyte-erythrocyte progenitor (MEP)) (**j**). Data are mean \pm s.e.m. *P* values

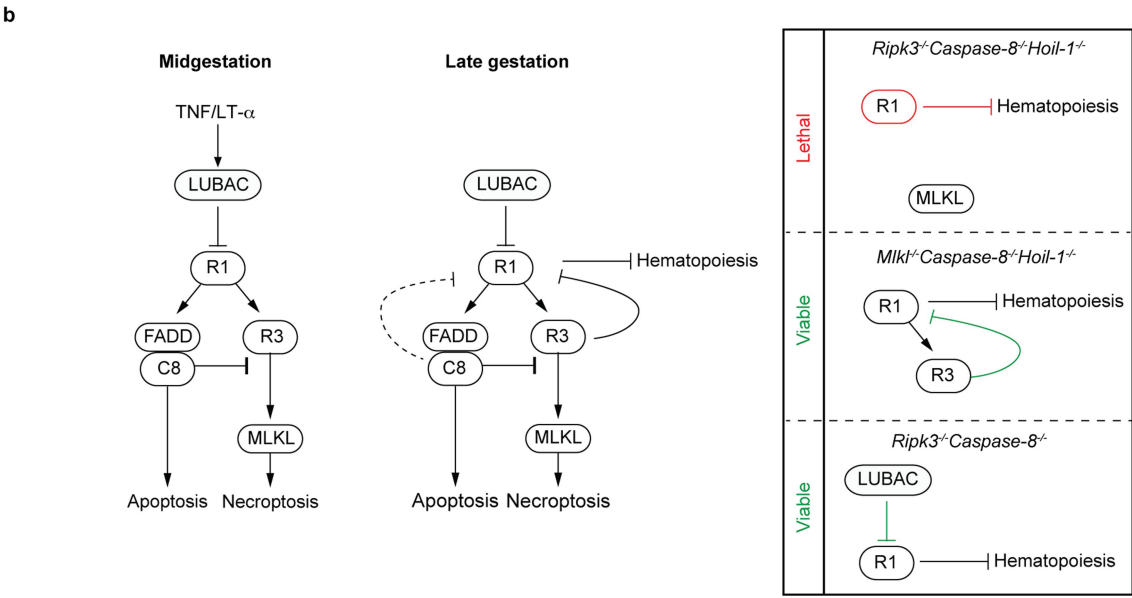
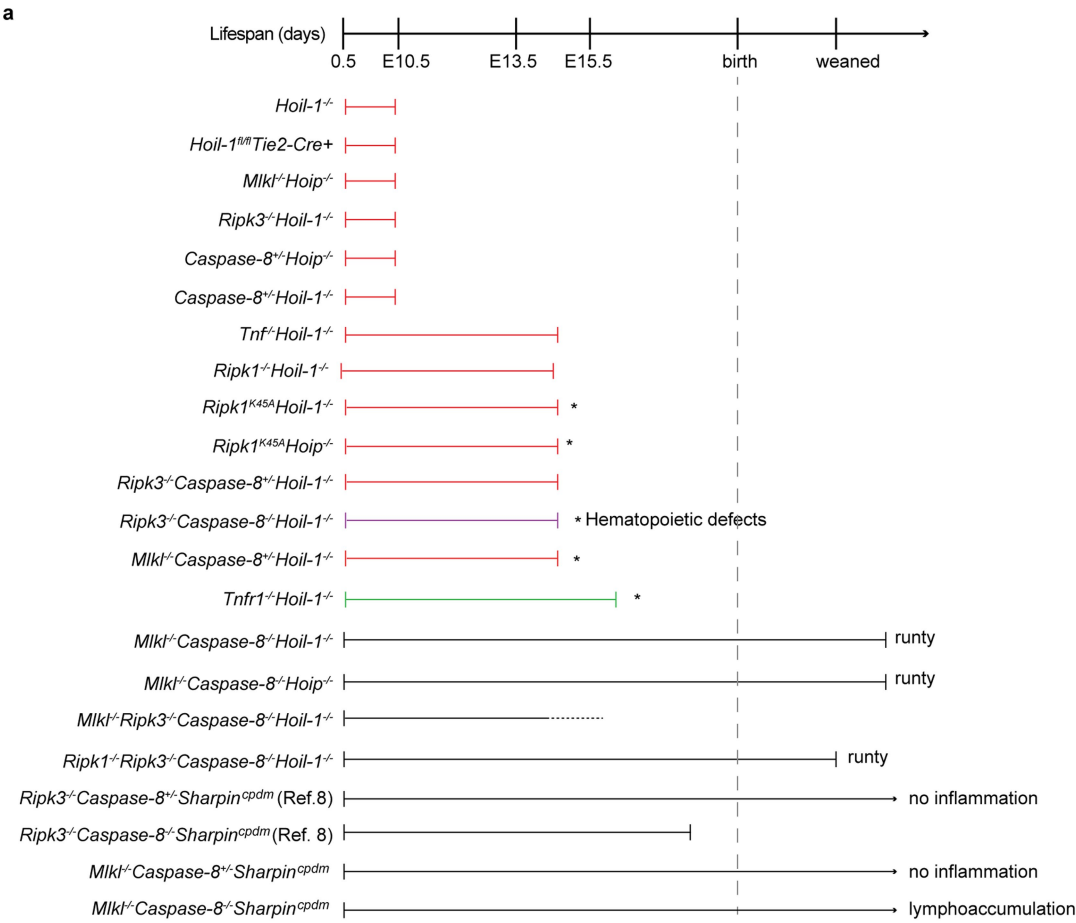
from unpaired two-tailed *t*-tests are shown. **g**, **i**, Percentages of mature CD45⁺ leucocytes, GR-1⁺ and F4-80⁺ cells (**g**) and CMP, GMP and MEP cells (**i**). Data are mean \pm s.e.m. *P* values from unpaired two-tailed *t*-tests are shown. **l**, Differentiation of E13.5 fetal liver (c-KIT⁺) progenitors into CFU-granulocytes and macrophages (GM), BFU-E and/or CFU-granulocyte, erythroid, macrophage, megakaryocyte (GEMM). Mean \pm s.e.m. ($n = 2$ fetal livers). **m**, Micrographs of differentiated macrophages ($n = 3$ *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-}, $n = 5$ *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and $n = 4$ *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal livers) and percentage viability of macrophages from E13.5 fetal liver cell suspensions from embryos of the indicated genotypes in the presence or absence of the indicated inhibitors. Data are mean \pm s.e.m. ($n = 3$ *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-}, $n = 5$ *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{+/-} and $n = 4$ *Mkl1*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} fetal livers). *P* values from two-way ANOVA are shown. **o**, Microfocus computed tomography scan images of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoil-1*^{-/-} embryos showing maximum intensity projections, with windowing applied to highlight vasculature (high contrast). No anatomical defects that would explain destruction of red blood cells or poor distribution of blood to the peripheries were found ($n = 3$ embryos). In the left image, yellow star denotes distal aorta, green star denotes umbilical vessels, and red star indicates descending thoracic aorta. In the right image, yellow star denotes carotid artery, red star denotes descending thoracic aorta, white star denotes ductus arteriosus, and blue star denotes ascending thoracic aorta. **p**, Representative FACS profile of a pool of three E11.5 dorsal aortas, containing the AGM region, per indicated genotype and quantification. This experiment was performed once with three embryos per genotype.



Extended Data Fig. 9 | Concomitant deletion of RIPK1 prevents embryonic lethality of *Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{-/-} mice.

a, Kaplan-Meier plot of mouse survival ($n = 17$ for *Ripk1*^{-/-}*Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{-/-} and $n = 2$ for *Ripk1*^{+/+}*Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{-/-} mice). **b**, Quantification of genotypes of animals obtained from intercrosses of *Ripk1*^{+/+}*Hoip-1*^{+/+} mice. For simplicity not all possible genotypes are represented. **c**, Percentage viability of macrophages from E13.5 fetal liver cell suspensions from embryos of the indicated genotypes. Data are mean \pm s.e.m. ($n = 5$ fetal livers/genotype). Statistical significance was determined with unpaired two-tailed *t*-tests. **d**, Cytokine arrays from *Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{+/+} and *Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{-/-}

embryos (left) and table listing the altered cytokines (right). Red squares highlight the differences ($n = 1$ for each genotype). For gel source data, see Supplementary Fig. 1. **e**, Cytokine analysis in homogenates from embryos of the indicated genotypes. Data are mean \pm s.e.m. ($n = 3$ embryos per genotype). *P* values from one-way ANOVA are reported. **f**, Representative images of E16.5 embryos from control mothers or mothers fed with the RIPK1 kinase inhibitor GSK'457A from mating and throughout gestation (embryos treated with GSK'457A $n = 5$ for *Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{+/+} and $n = 7$ *Ripk3*^{-/-}*Casp8*^{+/+}*Hoip-1*^{-/-} and $n = 3$ for *Ripk3*^{-/-}*Casp8*^{-/-}*Hoip-1*^{-/-}). Scale bar, 5 mm.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Schematic representation of findings in this study. a, Diagram indicating extent of viability and phenotypes of single, double, triple and quadruple knockout mice. Red lines indicate cell death and loss of yolk sac vascularization phenotype. Green line indicates mild cell death phenotype without loss of yolk sac vascularization. Asterisk indicates that heart defects were observed. **b,** Proposed model of LUBAC function during embryogenesis. At mid-gestation (left), LUBAC maintains vascular tissue integrity by preventing aberrant TNF/LT- α -mediated caspase-8- and RIPK3/MLKL-induced cell death. At late gestation, LUBAC is required not only to prevent aberrant cell death but also to prevent severe defects in haematopoiesis that are driven by RIPK1 but can be prevented by RIPK3 (middle). Genetic ablation of

LUBAC and of different components of the cell death machinery indicates that (right): (1) in the absence of LUBAC, caspase-8 and RIPK3, RIPK1 provokes lethality, probably by depleting multipotent progenitors in the haematopoietic compartment; (2) in the absence of caspase-8 and MLKL, cell death induced by loss of LUBAC is prevented and RIPK3 is present to exert its protective role on fetal haematopoiesis by precluding aberrant RIPK1 signalling; and (3) in the absence of caspase-8 and RIPK3, the presence of LUBAC is sufficient to prevent RIPK1 from causing severe defects in haematopoiesis and lethality since *Ripk3*^{-/-}*Casp8*^{-/-} mice are viable^{14,15,24}. This indicates that RIPK3 and LUBAC can compensate for each other to block aberrant RIPK1 signalling.

Architecture of an HIV-1 reverse transcriptase initiation complex

Kevin P. Larsen^{1,2,4}, Yamuna Kalyani Mathiharan^{3,4}, Kalli Kappel¹, Aaron T. Coey^{1,2}, Dong-Hua Chen², Daniel Barrero², Lauren Madigan², Joseph D. Puglisi², Georgios Skiniotis^{2,3} & Elisabetta Viani Puglisi^{2*}

Reverse transcription of the HIV-1 RNA genome into double-stranded DNA is a central step in viral infection¹ and a common target of antiretroviral drugs². The reaction is catalysed by viral reverse transcriptase (RT)^{3,4} that is packaged in an infectious virion with two copies of viral genomic RNA⁵ each bound to host lysine 3 transfer RNA (tRNA^{Lys}₃), which acts as a primer for initiation of reverse transcription^{6,7}. Upon viral entry into cells, initiation is slow and non-processive compared to elongation^{8,9}. Despite extensive efforts, the structural basis of RT function during initiation has remained a mystery. Here we use cryo-electron microscopy to determine a three-dimensional structure of an HIV-1 RT initiation complex. In our structure, RT is in an inactive polymerase conformation with open fingers and thumb and with the nucleic acid primer-template complex shifted away from the active site. The primer binding site (PBS) helix formed between tRNA^{Lys}₃ and HIV-1 RNA lies in the cleft of RT and is extended by additional pairing interactions. The 5' end of the tRNA refolds and stacks on the PBS to create a long helical structure, while the remaining viral RNA forms two helical stems positioned above the RT active site, with a linker that connects these helices to the RNase H region of the PBS. Our results illustrate how RNA structure in the initiation complex alters RT conformation to decrease activity, highlighting a potential target for drug action.

During the initiation phase of reverse transcription, RT must bind productively to the viral RNA–tRNA^{Lys}₃ complex and then navigate a highly-structured 5' region of the HIV-1 genome¹⁰. Critical elements within the viral RNA and host tRNA that are necessary for efficient initiation have been identified^{9,11–18}. RT pauses at discrete locations, is generally slowed during initiation compared to elongation^{8–10}, and can bind the viral RNA–tRNA^{Lys}₃ primer site in different orientations¹⁹. A rich body of structural data on RT, a heterodimer of p51 and p66 subunits, have shown how its polymerase and RNase H domains interact with DNA–DNA and DNA–RNA duplexes in the absence and presence of antiviral drugs². Lacking, however, are structures that reflect initiation, showing how RT binds to a large bimolecular viral RNA–tRNA^{Lys}₃ complex.

We used cryo-electron microscopy (cryo-EM) complemented by biochemical and biophysical experiments to determine the molecular architecture of an HIV-1 reverse transcriptase initiation complex (RTIC). The RTIC was formed using a 101-nucleotide fragment of HIV-1 genomic RNA (vRNA) that encompasses the primer binding site and additional RNA elements required for efficient initiation of reverse transcription (Fig. 1a). A binary vRNA–tRNA complex was formed with human tRNA^{Lys}₃ that contained a specific cross-linkable nucleotide ('convertible G') at position 71. The RTIC is kinetically labile, undergoing rapid RT dissociation from the tRNA–vRNA complex^{8,9} with several distinct binding orientations¹⁹. To stabilize the RTIC for structural characterization, the vRNA–tRNA complex was specifically cross-linked to RT containing a Q258C mutation in the p66 subunit (Fig. 1b), which interacts in the minor groove of RT–nucleic

acid complexes^{2,20,21}. After extending the tRNA primer by one dideoxynucleotide to achieve the highest crosslinking efficiency, we generated the cross-linked vRNA–tRNA–RT ternary complex and purified it from free RT and RNA (Fig. 1c, Extended Data Fig. 1a–c). Crosslinking did not affect the global activity of the RTIC. The final cross-linked HIV-1 RTIC had equivalent total activity in incorporation of the next dNTP as an un-crosslinked initiation complex, with rates that are only threefold slower, and is strongly inhibited by nevirapine, a non-nucleoside RT inhibitor (NNRTI) that works through conformational modulation of RT²² (Extended Data Fig. 1d–h). The RTIC studied structurally here thus represents an active functional state of reverse transcription initiation.

We first assessed the quality of the RTIC sample by negative stain electron microscopy²³, which confirmed a homogenous RT–RNA complex (Extended Data Fig. 2a). Upon cryo-EM preparation, however, the complex dissociated. This problem was alleviated by addition of beta-octyl glucoside, which resulted in monodisperse single particles that we visualized by cryo-EM (Extended Data Fig. 2b, c). Two-dimensional class averages of RTIC clearly showed the RT core as well as protruding RNA densities (Extended Data Fig. 2d). Three-dimensional classification of particle projections revealed substantial conformational variability in the apex of RNA densities (Extended Data Fig. 3a). Owing to this segmental flexibility, we obtained a low-resolution (8.0 Å) reconstruction that best describes the global architecture of RTIC (Fig. 2a, Extended Data Fig. 3b). This EM-density map, encompassing protein and all RNA regions, was of sufficient quality to visualize the tRNA and vRNA, thereby enabling us to position approximately the RNA structures located outside the RT binding cleft. In addition, we obtained a 4.5 Å map by masking out the dynamic peripheral RNA elements and focusing the particle classification and structure refinement on the RT, primer binding site (PBS) helix in the cleft, and additional helical tRNA density (Fig. 3, Extended Data Fig. 3). This higher-resolution map allowed us to describe the conformation of RT and the RNA inside the binding cleft (Fig. 3, Extended Data Fig. 4). An independent 8.2 Å cryo-EM reconstruction of the RTIC was determined in low salt and Mg²⁺ and revealed a very similar global conformation for the complex (Extended Data Fig. 5), suggesting that the RTIC architecture has limited salt dependence. Models were constructed using the 8.0 Å map to define the global RTIC architecture and the 4.5 Å map to define the structural features of the RTIC core and active site. While the 4.5 Å map provided sufficient resolution to orient the PBS helix of the RTIC, the orientation of the peripheral RNA helical elements of the vRNA and tRNA into the 8.0 Å map was more subjective and relied on iterative Rosetta²⁴ modelling using an accepted secondary structure from past biochemical and biophysical data^{14,25} (see Methods).

The overall RTIC structure shows the RT core with RNA double-helical density within the binding cleft that spans from the active site to the RNase H domain. The helical RNA in the cleft corresponds to the HIV-1 PBS helix formed between nucleotides (nts) 59–76 in the tRNA

¹Program in Biophysics, Stanford University, Stanford, CA, USA. ²Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA, USA. ⁴These authors contributed equally: Kevin P. Larsen, Yamuna Kalyani Mathiharan. *e-mail: epuglisi@stanford.edu

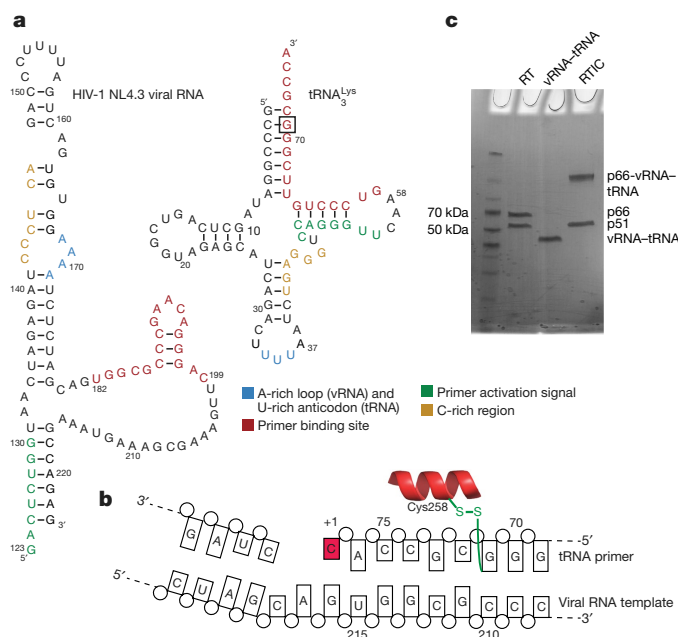


Fig. 1 | RTIC constructs and purification. **a**, HIV-1 viral RNA (NL4.3) and tRNA^{Lys} sequences used in RTIC formation. The viral RNA includes sequences complementary to the tRNA primer (coloured). Interactions between the regions are reported to be involved in the regulation of initiation. **b**, The crosslinking scheme used for purification of the RTIC. The N2-cystamine-dG was placed at position 71 of the tRNA primer. After extending the primer by one nucleotide, a ddCTP (red), a disulfide bond forms between G71 and mutated C258 on the RT p66 thumb subdomain. **c**, Non-reducing SDS-PAGE gel of the free vRNA-tRNA, RT, and crosslinked RTIC. RT runs as two bands corresponding to the two subunits. The annealed vRNA-tRNA complex runs as a single band on the gel. The purified RTIC runs as two bands corresponding to the p51 subunit and the crosslinked p66-vRNA-tRNA. Gel analysis was performed on all samples used in the manuscript (>10) and consistently exhibited similar results.

and 182–199 in the vRNA, with the addition of one ddCTP nucleotide needed to elongate the complex and allow efficient RT-RNA cross linking (Fig. 2c). The helical density for this +1 extended PBS helix is further extended near the RNase H domain by formation of an additional four base pairs, probably between complementary tRNA nts 55–58 and viral RNA nts 200–203. The nucleotide identities of positions 201–203 are highly conserved among recorded HIV-1 sequences²⁶ (70% for 201 and >96% for 202/203), suggesting that this is a common structural feature. In the 8.0 Å global map, a long continuous helical RNA density is observed to extend away from the RNase H domain (Fig. 2b). Accordingly, we propose that the 5' end of the tRNA (nts 1–54) refolds to form a secondary structure with a contiguous helix (Fig. 2c). Specifically, the D and anticodon stems from nts 10–44 rearrange to form a continuous helical structure, which fits the observed density far better than the three-way junction^{27,28} observed in the free initiation complex (Extended Data Fig. 6).

The helical refolded tRNA domain is connected by a single-stranded connection loop to a 7-bp helix (H1) involving the 5' (nts 125–131) and 3' (nts 217–223) termini of the viral RNA construct (Fig. 2b, c) and containing the conserved primer activation signal (PAS) sequence. H1 and the connection loop form a bridge between the RNA located in the RNase H domain and that located near the active site of RT. A three-way RNA junction is formed by the PBS, H1 and a second helical stem loop (H2) comprising nts 134–178 of HIV-1 viral RNA. Density consistent with single-stranded RNA connects H2 to the PBS in the active site. The relative strength (indicative of stability) of the EM density for H1, the connection loop, and the apical regions of H2 differs among several of our low-resolution classes, as do their orientations with respect to the base of H2 and the PBS (Extended Data Fig. 7a, b). For classes that

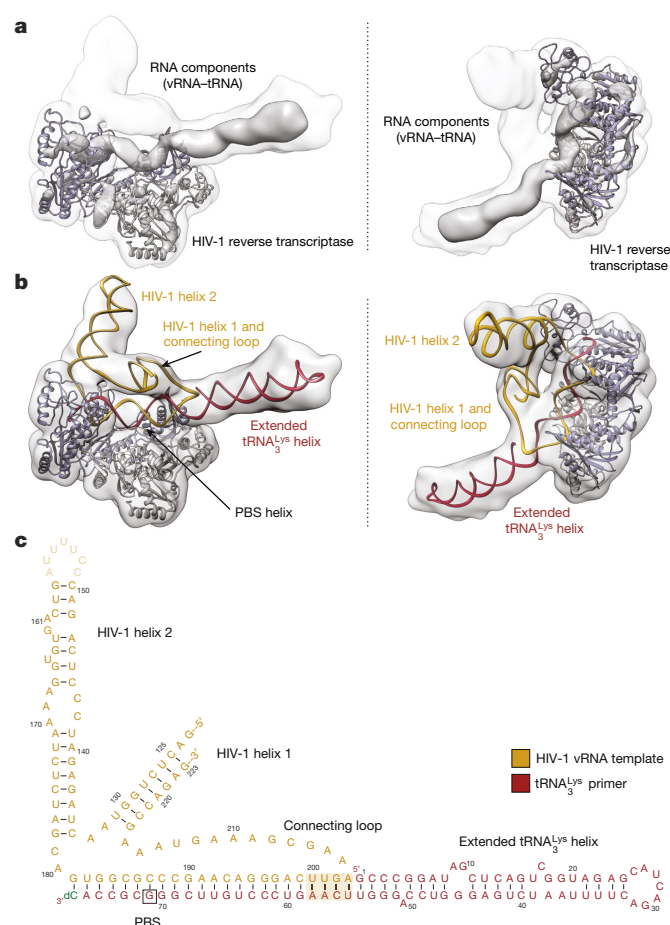


Fig. 2 | Global architecture of the RTIC. **a**, Unmasked 3D reconstruction of the entire RTIC at a global resolution of 8.0 Å. A model of RT (p66, purple; p51, grey) is docked into the map. A low-threshold map has been overlaid into the density to illustrate the helical nature of the bound RNA duplex. **b**, Global model of the RTIC with the vRNA-tRNA components. The electron micrograph density accounts for the majority of vRNA and tRNA structure. Density corresponding to the upper HIV helix 2 stem loop is missing, suggesting that it is partially disordered. **c**, Proposed secondary structure of the vRNA-tRNA bound within the RTIC. The majority of vRNA helices are well accounted for in the density, with the exception of the apical portions of helix 2 (faded). Additional base pairs (boxed) between the vRNA and the tRNA, which extend the PBS helix, are consistent with the continuous helix that spans the RT binding cleft. The tRNA has refolded and adopted an extended helical conformation.

contain strong density of these RNA features, similar models fit these maps by treating the helical RNA elements as rigid units around flexible junction regions (Extended Data Fig. 7c). The presence of helix H1 was confirmed by single-molecule Förster resonance energy transfer (FRET) experiments, in which Cy3 dye was attached to the 5' phosphate of the vRNA and a Cy5-labelled oligonucleotide was hybridized to an extension on the 3' end (Extended Data Fig. 8a). In this experiment, observation of a high FRET state would indicate H1 formation. In the buffer conditions used for cryo-EM imaging, we find that more than 95% of RTIC molecules are in a stable, high-FRET state, indicating that H1 forms for a surface-immobilized RTIC at room temperature (Extended Data Fig. 8b, c).

Although the RTIC is active in the addition of the next dNTP (Extended Data Fig. 1d, e, h), the complex adopts an inactive conformation in which the position of the tRNA primer terminus within the palm subdomain is shifted approximately 13 Å away from the active site of RT, reminiscent of nucleic acid-RT complexes bound with an NNRTI²² (Fig. 4a, Extended Data Fig. 9b). As observed in RT structures with bound NNRTI, the primer grip (B12–B13–B14 sheet) is displaced towards the 3' terminus of the primer strand²². The PBS helix is not

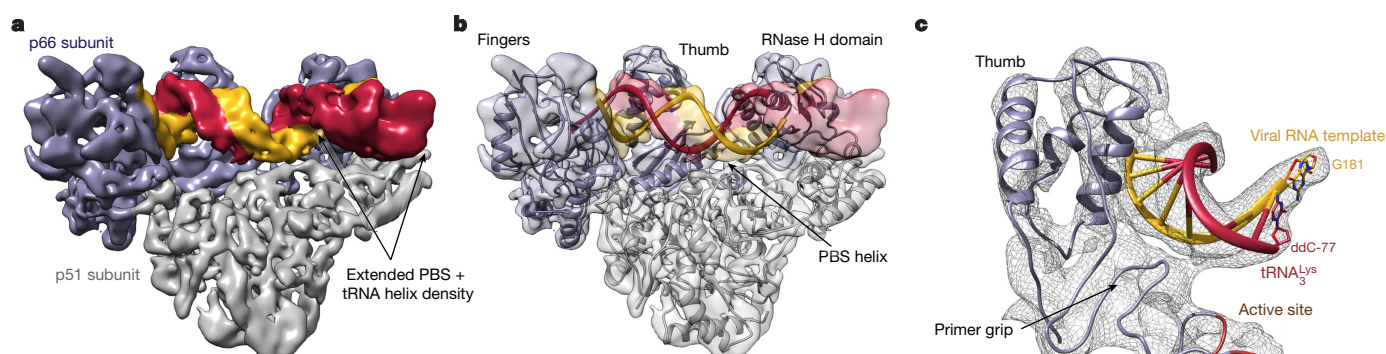


Fig. 3 | Structure of the RTIC core. **a**, Cryo-EM density map of the RTIC core at 4.5 Å resolution. The helical PBS density is regular until it reaches the RNase H active site. Extra helical density for an additional four base pairs between the vRNA and tRNA is present. Density corresponding to the tRNA is located just outside the RNase H domain. The junction between the extended PBS and tRNA helix is distorted, possibly owing to masking or flexibility in this region. This portion of the RTIC is

helical in the global map (Fig. 2). **b**, Representative model of the RTIC core that accounts for RT and the entire +1 extended PBS helix. There is helical density that can accommodate additional base pairs between the vRNA and tRNA, but we have not included it in this model. **c**, Fit of the polymerase active site region in the 4.5 Å map. The RNA and protein backbone are well modelled in this region of the map.

translocated, with the dC77–G181 pair in the nucleotide acceptor site (N-site) (Fig. 4). The PBS helix is also lifted about 6.0 Å away from the palm and connection domains. The path of the viral RNA template and the base of H2 come into close contact with the residues in the fingers domain. The fingers domain of RT adopts a semi-open conformation similar to that of RT structures bound to nucleic acids that lack an incoming nucleotide^{2,22} (Fig. 4a, Extended Data Fig. 9). On the basis of previous mechanistic studies of RT enzymology^{8,29}, we conclude that the RTIC here is blocked in a pre-translocation conformation for the primer–template complex (Fig. 4c). Unlike NNRTI-bound complexes, the RTIC is functional and can incorporate the next dNTP, suggesting that there is conformational plasticity within the RT active site. Although RT contacts the RNA substrate using similar domains as in previously determined RT–nucleic acid complexes^{2,20,22}, the extent of these interactions appears different. The thumb and RNase H domains make the vast majority of observed RNA contact in the RTIC, with a substantial loss of potential interactions in the palm and connection subdomains, consistent with decreased RT–RNA affinity in the initiation complex. The loss of RT–RNA contacts in the palm subdomain²⁰ arises from displacement of the tRNA primer terminus away from the active site (Fig. 4a). Although the RTIC structure is not at sufficient resolution to identify specific protein–RNA contacts, there appear to be additional RT–RNA interactions involving the fingers domain with the vRNA template–strand and H2. The sterically bulky vRNA helices immediately adjacent to the fingers region form a wedge that hinders proper accommodation of the PBS into the cleft and leads to loss of RT–RNA contacts in the cleft and displacement of tRNA 3′ end (Fig. 4b); this is likely to inhibit translocation of the PBS helix to enable efficient and rapid incorporation of the next dNTP.

The architecture of the vRNA–tRNA complex in the initiation complex explains previous experimental results on the role of RNA in initiation^{8–10}. The observed RNA conformation is consistent with chemical probing and enzymatic mapping on similar binary vRNA–tRNA and ternary complexes, which were previously interpreted in terms of tRNA–viral RNA pairings^{13,14,25}. No additional interactions between the vRNA and tRNA occur beyond the extended PBS helix at the +1 stage of initiation (Fig. 2b), consistent with biochemical results on similar HIV-1 subtype-B sequences^{13,14}. Notably absent is any PAS–anti-PAS interaction between HIV-1 nts 123–130 and tRNA nts 48–55 (Fig. 1a), which has been implicated in RT initiation and shown to form dynamically in the absence of RT^{11,12,27,30}. The formation and positions of vRNA H1 and H2 are consistent with their proposed function as barriers during initiation^{13,14,19}. The conserved connection loop, bridging RNA within the RNase H domain back to H1, may help to position the vRNA helices in the proper orientation for binding of RT to the tRNA 3′ terminus. The HIV-1 genomic RNA from the MAL

isolate, commonly used in past initiation studies, maintains many of these sequence elements with an added 23-nt insertion in the connection loop that may engage in additional interactions. Both H1 and H2 are required for efficient initiation of reverse transcription, and their displacement and unfolding are required for reverse transcription to proceed^{9,11,15,17}; melting of H2 during initiation occurs after addition of the sixth nucleotide¹⁹.

tRNA^{Lys}₃ in the RTIC forms an elongated helical structure compatible with an alternative predicted fold³¹ that involves an extended PBS structure stabilized by RT. This conformation is consistent with the presence of modified nucleotides in tRNA^{Lys}₃ (Extended Data Fig. 6d), and is favoured by the extended stacking and RT contacts around the RNase H domain. The RNA fold in the RTIC is likely to sequester important sequences for vRNA–tRNA interactions, such as the PAS–anti-PAS and A-rich loop–anticodon sequence interactions^{9,11–18}, which may subsequently form as RNAs rearrange in response to RT extension during initiation. The RNA tertiary conformation within the RTIC is clearly dynamic, as shown by published single-molecule data^{19,27,30} and suggested by our cryo-EM data. We observe several conformations with variable orientations and density for the extended tRNA helix and three-way junction of H1, H2 and PBS (Extended Data Fig. 7). Such plasticity is likely to be essential for the RTIC to proceed to elongation.

Our results suggest a model of RT initiation in which RNA structure regulates RT activity. tRNA^{Lys}₃ and vRNA form a dynamic RNA complex, in which the tRNA refolds to form a metastable conformation of its 5′ region. The ability to refold in this way could explain the use of tRNA^{Lys}₃ in HIV-1 initiation. Although RT contacts the PBS in the cleft, the disrupted palm subdomain contacts between the PBS and RT explain the poor affinity of RT for the vRNA–tRNA complex. Within the framework of the standard dNTP incorporation mechanism²⁹, RT in this +1 initiation complex adopts a pre-translocation conformation with an open active site and improper positioning of nucleic acid for catalysis (Fig. 4c). The vRNA helices, whose orientation hinders productive binding to RT, must be displaced and/or unfolded for the tRNA primer terminus to reposition within the active site such that the RT fingers can clamp down on an incoming nucleotide (Fig. 4b). The dissociation of RT during initiation is rapid, and competes with forward polymerization reactions^{8,9}. RT may dissociate and rebound to the vRNA–tRNA to reposition the primer terminus into the active site. In this pathway, RT rebinding could facilitate melting of downstream RNA structures that hinder translocation. The necessity for these rearrangements during early stages of initiation is likely to explain the low processivity of initiation and the observed pauses that control the start of HIV-1 replication^{8,9}. The single-stranded, A-rich connection loop bridging the 3′ end of the vRNA PBS to H2 may position the vRNA

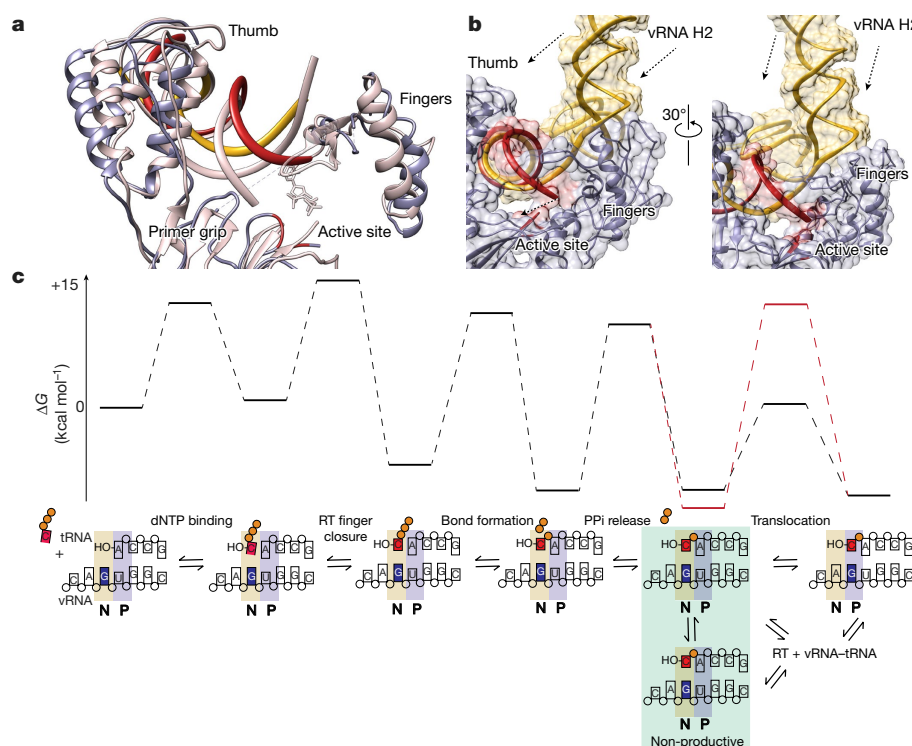


Fig. 4 | The +1 RTIC adopts an inactive conformation. **a**, Comparison of RTIC primer (red) and template (yellow) strands with RT-dsDNA complex (pink, 1RTD²³) that has the 3' primer terminus located in the P-site with the N-site occupied by a nucleotide. The PBS helix of the RTIC must be translocated and shifted in order to reposition into a P-site conformation. The thumb is in an open conformation and the primer grip has shifted compared to the active structure. The fingers are in a semi-open conformation²². **b**, vRNA structure outside the RT active site may prevent proper translocation of RNA substrate during initiation. Two views of the RNA (vRNA, yellow; tRNA, red) near the active site of RT. The arrows indicate the direction in which the RNA must move in order for the PBS to reposition into the active site. The global structure

here is represented in surface mode to highlight potential steric clashes. **c**, Reaction coordinate diagram and corresponding reaction scheme of HIV-1 reverse transcription on an RNA template using published values²⁹. The +1 RTIC appears to be trapped in a pre-translocation state (green box) following pyrophosphate (PPi) release. The increased energetic barrier for translocation (red) corresponds to the energetic requirement for melting base pairs in H2. We have included the possibility of a non-productive conformer at the pre-translocation step to acknowledge our inability to completely distinguish an on-pathway complex from one that is relaxed and off-pathway. There is also an alternative pathway in which the RTIC dissociates and reassociates to forgo translocation and reposition the primer terminus.

helices properly and allow conformational communication with the RT RNase H domain and refolded tRNA. As reverse transcription proceeds, structural rearrangements in vRNA and tRNA must occur to favour the transition to processive elongation. Thus, the initiation complex is likely to change progressively as initiation proceeds, and may be specifically vulnerable to inhibition by drugs. Higher-resolution structural views of these different states, and dynamics to link them together, will be needed to elucidate further the steps of initiation and underlying RNA conformations that regulate early steps in HIV-1 infection.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0055-9>.

Received: 18 September 2017; Accepted: 19 March 2018;

Published online 25 April 2018.

- Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. A detailed model of reverse transcription and tests of crucial aspects. *Cell* **18**, 93–100 (1979).
- Sarafianos, S. G. et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385**, 693–713 (2009).
- Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209–1211 (1970).
- Temin, H. M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211–1213 (1970).
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* **2**, 461–472 (2004).

- Huang, Y. et al. Incorporation of excess wild-type and mutant tRNA(3Lys) into human immunodeficiency virus type 1. *J. Virol.* **68**, 7676–7683 (1994).
- Ratner, L. et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**, 277–284 (1985).
- Lanchy, J. M., Ehresmann, C., Le Grice, S. F., Ehresmann, B. & Marquet, R. Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO J.* **15**, 7178–7187 (1996).
- Lanchy, J. M. et al. Contacts between reverse transcriptase and the primer strand govern the transition from initiation to elongation of HIV-1 reverse transcription. *J. Biol. Chem.* **273**, 24425–24432 (1998).
- Isel, C., Ehresmann, C. & Marquet, R. Initiation of HIV reverse transcription. *Viruses* **2**, 213–243 (2010).
- Beerens, N. & Berkhout, B. The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. *J. Virol.* **76**, 2329–2339 (2002).
- Beerens, N., Groot, F. & Berkhout, B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.* **276**, 31247–31256 (2001).
- Goldschmidt, V., Ehresmann, C., Ehresmann, B. & Marquet, R. Does the HIV-1 primer activation signal interact with tRNA_{3Lys} during the initiation of reverse transcription? *Nucleic Acids Res.* **31**, 850–859 (2003).
- Goldschmidt, V. et al. Structural variability of the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* **279**, 35923–35931 (2004).
- Goldschmidt, V. et al. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. *J. Biol. Chem.* **277**, 43233–43242 (2002).
- Isel, C. et al. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* **18**, 1038–1048 (1999).
- Iwatani, Y., Rosen, A. E., Guo, J., Musier-Forsyth, K. & Levin, J. G. Efficient initiation of HIV-1 reverse transcription *in vitro*. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. *J. Biol. Chem.* **278**, 14185–14195 (2003).
- Liang, C. et al. The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J. Virol.* **71**, 5750–5757 (1997).

19. Liu, S., Harada, B. T., Miller, J. T., Le Grice, S. F. & Zhuang, X. Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat. Struct. Mol. Biol.* **17**, 1453–1460 (2010).
20. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
21. Jacobo-Molina, A. et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA* **90**, 6320–6324 (1993).
22. Das, K., Martinez, S. E., Bauman, J. D. & Arnold, E. HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat. Struct. Mol. Biol.* **19**, 253–259 (2012).
23. Peisley, A. & Skiniotis, G. 2D projection analysis of GPCR complexes by negative stain electron microscopy. *Methods Mol. Biol.* **1335**, 29–38 (2015).
24. Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* **7**, 291–294 (2010).
25. Watts, J. M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716 (2009).
26. Foley, B. et al. *HIV Sequence Compendium 2013* (Los Alamos National Laboratory, Los Alamos, 2013).
27. Coey, A., Larsen, K., Puglisi, J. D. & Viani Puglisi, E. Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *RNA* **22**, 1689–1698 (2016).
28. Puglisi, E. V. & Puglisi, J. D. Secondary structure of the HIV reverse transcription initiation complex by NMR. *J. Mol. Biol.* **410**, 863–874 (2011).
29. Li, A., Gong, S. & Johnson, K. A. Rate-limiting pyrophosphate release by HIV reverse transcriptase improves fidelity. *J. Biol. Chem.* **291**, 26554–26565 (2016).
30. Beerens, N. et al. Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA* **19**, 517–526 (2013).
31. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

Acknowledgements We thank A. Frost and L. Stryer for suggesting beta-octyl glucoside as an additive for cryo-EM, R. Kornberg, M. Levitt, P. Geiduschek and W. Sundquist for reading the manuscript, M. Levitt for discussion of alternative tRNA folds and general support, D. Herschlag for discussions, and N. R. Latorraca for discussions and assistance with the Sherlock cluster. Supported

by National Institutes of Health grant GM082545 to E.V.P., T32-GM008294 (Molecular Biophysics Training Program) to K.P.L., A.T.C. and K.K., National Science Foundation Graduate Research Fellowship Program (DGE-114747) to A.T.C. and K.K., and Gabilan Stanford Graduate Fellowship to K.K. We thank Stanford University and the Stanford Research Computing Center for providing the Sherlock cluster resources. Additional calculations were performed on the Stanford BioX3 cluster, supported by NIH Shared Instrumentation Grant 1S1ORR02664701.

Reviewer information *Nature* thanks N. Sluis-Cremer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.P.L., Y.K.M. and D.-H.C. acquired preliminary cryo-EM data and performed initial cryo-EM map calculations. Y.K.M. acquired cryo-EM data and obtained the 3D reconstructions shown in the main manuscript. K.P.L. acquired Mg²⁺ cryo-EM data and performed corresponding cryo-EM map calculations. A.T.C. purified the vRNA used for single-molecule experimentation and performed the single-molecule experiments. K.P.L., D.B. and L.M. performed all vRNA and RT sample preparations. K.P.L. performed all α -³²P-dTTP incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. K.P.L. and Y.K.M. performed final RTIC model building and refinement. K.P.L., Y.K.M., G.S., J.D.P. and E.V.P. interpreted the data. K.P.L. and E.V.P. wrote the manuscript with input from J.D.P., K.K., Y.K.M. and G.S.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0055-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0055-9>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to E.V.P.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Sample preparation. HIV-1 vRNA constructs were prepared by in vitro transcription with T7 RNA polymerase as previously described^{27,28}. Transcripts were denatured in 8 M urea and purified on a sequencing PAGE gel. Gel extraction was performed using 0.3 M ammonium acetate. Following ethanol precipitation, the RNA was dissolved in 10 mM Bis-Tris propane, pH 7.0, 10 mM NaCl and stored at -20°C . The crosslinkable tRNA^{Lys}₃ construct was purchased from TriLink Biotechnologies. The crosslinkable RNA primer was chemically synthesized, PAGE purified, and analysed by denaturing PAGE and mass spectrometry. During synthesis, an N²-cystamine-2'-deoxyguanosine was placed at the 71 position for crosslinking purposes.

vRNA-tRNA complexes were formed by mixing the vRNA and tRNA in a 1:1 molar ratio at 1 μM each in 10 mM Bis-Tris propane, pH 7.0, 10 mM NaCl. The mixture was heated to 90°C and slow cooled to room temperature. The vRNA-tRNA complex was purified away from higher order and unannealed monomer species using a Superdex 200 (26/60) gel filtration column with 10 mM Bis-Tris propane, pH 7.0, 100 mM NaCl. The presence of a single species was confirmed with native PAGE and samples were concentrated on a Vivaspinn 20 10,000 MWCO concentrator. Samples were stored at -20°C and exhibited minimal aggregation over time.

HIV-1 RT was expressed in *Escherichia coli* strain BL21(DE3). Two expression vectors, one containing p66 and ampicillin resistance and the other containing p51 and kanamycin resistance, were constructed. The C terminus of p66 contains an unstructured linker and a six-histidine tag. A cysteine mutation for crosslinking was introduced into helix H of p66 (Q258C)²⁰. The protein used in this study also had the C280S mutation, introduced in prior structural work, and the E478Q mutation, introduced to eliminate RNase H activity as RT has been shown to cleave dsRNA when stalled for long periods^{20,32}. Cell pellets were lysed through sonication and the enzyme was purified by gravity Ni-nitrilotriacetic acid (Ni-NTA) affinity chromatography, followed by size-exclusion chromatography using a Superdex 200 (26/600). The His₆ tag was cleaved by thrombin digestion overnight. The cleaved protein was re-applied to a Ni-NTA column to remove protein with an uncleaved His₆ tag. This was followed by an additional final size-exclusion chromatography step. The protein was stored at 4°C in 300 mM NaCl, 50 mM Tris, pH 8.0, 5 mM β -met.

The RTIC was prepared by mixing RT and vRNA-tRNA complex at 2 and 1 μM , respectively, in a buffer containing 25 mM NaCl, 25 mM KCl, 5 mM MgCl₂, 50 mM Tris, pH 7.5, 100 μM ddCTP (or dCTP if used for +2 incorporation assays). The mixture was allowed to crosslink overnight at room temperature. The complex was purified by anion-exchange chromatography with a linear gradient. This was followed by a size-exclusion chromatography step to remove any higher-molecular-weight aggregates. The purity and homogeneity of the final complex were assessed by SDS-PAGE (under non-reducing conditions) and size-exclusion chromatography (Extended Data Fig. 1).

Amino-GMP-labelled viral RNA for single-molecule experimentation was transcribed as previously described but with nucleotide concentrations of 1 mM ATP, CTP and UTP and 0.5 mM GMP. The vRNA sequence is identical to that used in the cryo-EM experiments, but contains an additional unstructured sequence on the 3' end for immobilization and oligonucleotide hybridization purposes and an additional GGU on the 5' end for labelling purposes. 5'-Amino-G-monophosphate (GMP), purchased from TriLink Biotechnologies, was added to the reaction at a final concentration of 1 mM. The reaction was incubated at 37°C for 4 h. 5'-Amino-GMP-labelled RNA was purified by phenol/chloroform extraction followed by a 10DG (Bio-Rad) desalting column in 10 mM Bis-Tris (pH 7.0), 75 mM NaCl. The RNA was then separated from template DNA and free NTPs by size-exclusion chromatography (ENRICH SEC 650 10 \times 300) in 100 mM sodium phosphate buffer (pH 8.2), 75 mM NaCl. Purified amino-GMP-labelled RNA was concentrated to 1 μM and labelled using NHS chemistry with 1,000-fold excess cyanine dye (Lumiprobe). Excess dye was removed by passage over a 10DG desalting column followed by size-exclusion chromatography (ENRICH SEC 650 10 \times 300) purification to buffer exchange the labelled vRNA. Labelling efficiency was calculated by measuring the absorbance values of the labelled species at both 260 nm (RNA absorbance) and 550 nm (Cy3 absorbance). These absorbance values were used to calculate the concentrations of the RNA and the Cy3 dye. Using the ratio between these two values, we estimate that our 5' labelling efficiency is approximately 70%.

Dye-labelled vRNA-tRNA complexes were heat-annealed and purified as previously described. The single-molecule RTIC complex was prepared as stated above, but with a several modifications. To simplify the purification, the his-tag was kept on the p66 subunit of RT. The RTIC was then applied to a Ni-NTA column and washed with 300 mM NaCl to remove the free vRNA-tRNA complexes. The RTIC was eluted from the column. Synthetic oligonucleotides with sequences 5'-GCGGGAGAUCAGGCAU(Am6)-cyanine5-3' and 5'-biotinCUAUCCCCU-AUCCdC-3' (Trilink) were annealed to the complex at 37°C for 5 min in tenfold molar excess. Excess oligonucleotides and free RT were rinsed away during TIRF

slide preparation. The above protocol was also performed for a dye-labelled vRNA-tRNA-only control, but skipped the RTIC complex formation, purification, and free RT rinse.

Single-molecule FRET experiments. Single-molecule FRET experiments were performed using a prism-based total internal reflection instrument with a diode-pumped solid-state 532-nm laser as previously described^{27,33–38}. This includes the use of an oxygen scavenging system (protocatechuate 3,4-dioxygenase (PCD) and β -carboxy-*cis*, *cis*-muconic acid (PCA)) and a triplet state quencher (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox)) to reduce aberrant dye behaviours. The laser power measured 50 mW at the prism. The fluorescence signal was recorded with an exposure of 100 ms per frame for 5 min at room temperature. FRET traces were manually analysed using home-written scripts in MATLAB (MathWorks)²⁷. This analysis began by using a colocalization script to select only spots that exhibited both Cy3 donor and Cy5 acceptor fluorescence (under donor-only excitation conditions). Such colocalization allowed us to eliminate partially labelled molecules. Next, FRET traces were manually inspected to eliminate cases with multiple single-dye photobleaching events (multiple molecules) or traces that exhibited poor dye photophysics (Extended Data Fig. 8d). After this manual inspection, the final dataset used for analysis included only traces in which both dyes exhibited clear single photobleaching events to ensure reliable data (Extended Data Fig. 8c). For our single-molecule experiment, 708 traces were selected through colocalization. After manual inspection and elimination of poor traces, 480 traces were used for the final analysis. In addition to the RTIC experiment, a control experiment using a dye-labelled vRNA-tRNA was performed to assess the FRET states without RT. We found that in the absence of cross-linked RT, a small population of low-FRET-state molecules with FRET efficiency 0.3 existed, but this state was not observed upon binding and cross-linking of RT (data not shown).

Negative-stain EM. We applied 3.5 μL of 0.1 μM RTIC sample onto glow-discharged carbon-coated grids, and blotted and stained them with 1% uranyl formate according to standard protocols²³. Negative-stained grids were imaged on an FEI Morgagni at 100 kV.

Cryo-EM data acquisition. RTIC complex in high monovalent salt buffer (300 mM NaCl, 10 mM Tris-HCl pH 8.0) containing 0.2–0.25% (w/v) beta-octyl glucoside (β -OG) was applied to glow-discharged holey carbon grids (Quantifoil R2/2, 200 mesh) and subsequently vitrified using a FEI Vitrobot. Frozen hydrated samples were imaged on an FEI Titan Krios at 300 kV with a Gatan K2 Summit direct detection camera in counting mode with 200 ms exposure per frame. Forty frames per micrograph were collected at a magnification of $29,000\times$, corresponding to 1 \AA per pixel at the specimen level. In total, 4,209 micrographs were collected at defocus values ranging from -1.3 to $-2.5\text{ }\mu\text{m}$. The movie frames were motion-corrected and dose-weighted by MotionCor2³⁹ and CTF parameters were estimated by CTFFIND4⁴⁰.

RTIC complex in low monovalent salt buffer and Mg²⁺ (75 mM NaCl, 2 mM MgCl₂, 10 mM Tris-HCl pH 8.0) containing 0.2% (w/v) β -OG was applied to glow-discharged holey carbon grids (EMS, 200 mesh, Copper) and subsequently vitrified using a Leica EM GP. Frozen hydrated samples were imaged on a Tecnai F20 at 200 kV with a Gatan K2 Summit direct detection camera in counting mode with 200 ms exposure per frame. Sixty frames per micrograph were collected at a magnification of $29,000\times$, which corresponds to 1.286 \AA per pixel at the specimen level. In total, 898 micrographs were collected at defocus values ranging from -2.0 to $-3.0\text{ }\mu\text{m}$ and a dose rate of 8.0 electrons per pixel per second. The micrograph movies were motion-corrected and dose-weighted as above, and CTF parameters were estimated by GCTF⁴¹.

Cryo-EM data processing. Cryo-EM data for the 8.0 and 4.5 \AA maps were processed using Relion^{42–44}. 765,688 particle projections were semi-automatically picked from the motion-corrected micrographs, and sorted through subsequent rounds of reference-free 2D classification. 444,374 particle projections belonging to classes with well-defined RT and RNA features were selected for further processing (Extended Data Fig. 3a). An initial 3D model was obtained using VIPER⁴⁵ based on the selected 2D classes, and used for 3D classification in Relion^{42,44} (Extended Data Fig. 3a). Because particle alignment was affected by the flexible protruding RNA, we used a mask and focused the alignment on RT and PBS alone. 167,906 particle projections sorted to 3D classes displaying all features of RT and PBS were selected for subsequent 3D classifications. To further improve the quality of RT/PBS core, one more round of 3D classification with finer angular sampling was executed; particles from two classes with well-defined secondary structure densities were combined and the 3D structure was refined to a resolution of 4.5 \AA . For global RTIC maps (including the flexible protruding RNA) a 3D classification without mask using the 167,906 particle projections subset was performed; eight classes obtained showed the tRNA and vRNA in various conformations. The class displaying most of the RNA protrusions was refined to a resolution of 8.0 \AA . The resolution reported is according to the 0.143 'gold standard' Fourier shell correlation (FSC) criterion (Extended Data Fig. 3b). The 4.5 and 8.0 \AA maps were corrected for the

modulation transfer function (MTF) of K2 direct detection camera at 300 kV and then sharpened using *B* factors of -250 and -200 \AA^2 , respectively, during the post-processing step (Extended Data Table 1). Local resolution was estimated using Relion (Extended Data Fig. 3c).

Cryo-EM data for the 8.2 \AA Mg^{2+} map were processed using Relion. 148,523 particle projections were semi-automatically picked from motion-corrected micrographs and sorted through subsequent rounds of reference-free 2D classification. 125,615 particle projections belonging to classes with well-defined RT and RNA features were selected for further processing (Extended Data Fig. 3e). An initial 3D model was obtained using EMAN2 based on selected 2D classes⁴⁶ and used for 3D classification. The resolution reported is according to the 0.143 'gold standard' FSC criterion (Extended Data Fig. 3f). The maps were corrected for the modulation transfer function (MTF) of K2 direct detection camera at 200 kV and then sharpened using a *B* factor of -200 during the post-processing step (Extended Data Table 1).

Model building and refinement. The crystal structure of RT bound to a DNA–DNA duplex, with nucleic acid substrate removed, was used as a starting model for RT²². After manually fitting the main-chain backbone of RT into its distinct density, four regions of EM density corresponding to RNA were apparent. The most notable region of density is the well-formed RNA helix in the cleft of RT, which corresponds to the PBS helix of the vRNA–tRNA complex. The complete model of the vRNA–tRNA complex was built piecewise, and iteratively, using the Rosetta FARFAR method²⁴. First, initial models of the first seven base pairs of the PBS helix (vRNA residues 181–187 and tRNA residues 71–77 with the extended dC, originally modelled as RNA for simplicity and later edited to a ddC in Coot⁴⁷) were built with FARFAR, then clustered. The centres of the ten most populated clusters were fit into the density using the colores tool in the Situs package⁴⁸. The resulting models were manually inspected and selected on the basis of fit to the density and proximity of G71 to C258 on RT (base and residue involved in crosslinking). This helix was extended out to nineteen base pairs (to include vRNA residues 187–199 and tRNA residues 59–70), models were again clustered, and the cluster centres were fit into the density. The model that best fit the 4.5 \AA cryo-EM density was selected. This RT–PBS model, called the RTIC core, was then refined using Phenix real-space-refinement⁴⁹ with secondary structure restraints in place for the RNA and protein. To further restrain the model during refinement, the N2-cystamine-deoxyguanosine was inserted into the tRNA sequence and a loose disulfide bond constraint with C258 was used during refinement (this was later reverted to a dG as there was no density for the linker atoms). The model was visually inspected and manually adjusted in Coot⁴⁷. Protein residues lacking EM density, the vast majority of which were located in the fingers and palm subdomains, were removed after comparison to prior models of RT (Extended Data Fig. 4). Owing to insufficient resolution, large regions of RT did not exhibit reasonable density for sidechains. Therefore the RT model was truncated to a main-chain backbone before final inspection and submission to the PDB. The geometry of the final refined model was validated using Molprobity⁵⁰. This refined RTIC core model (Fig. 3) served as the anchor point for orienting the vRNA and tRNA portions of the global RTIC model.

Two additional regions of RNA density were located near the fingers subdomain of RT (Fig. 2a). As the crosslinking method used to form the complex harnesses RT polymerase activity, these regions of density correspond to the template vRNA helices. We traced the vRNA template strand out of the active site, allowing us to confidently orient the base of vRNA H2. Models of vRNA H2 (residues 134–178) were built with FARFAR²⁴ based on the consensus secondary structure from past biochemical and biophysical data^{25,51}. These models were clustered and fit into the density.

After positioning vRNA H2, only one region of RNA density near the fingers remained unaccounted. This density, which was continuous with the H2 density, corresponds to vRNA H1 (residues 125–131 and 217–223). This density also connects to RNA located in the cleft of RT near the RNase H domain. This suggests that the connection loop may contact H1 and contribute to the density observed in this region. To confirm the presence of H1 in the RTIC, single-molecule experiments were performed in which a FRET pair was placed on the 5' and 3' ends of the helix. We find that in our imaging conditions, 85–95% of molecules exist in a stable high FRET, consistent with H1 formation. vRNA helix 1 was modelled as an ideal A-form helix, then fit into the density using UCSF Chimera⁵². This initial fit was later refined during the global model building (described below). The connecting loop was partially built in Coot⁴⁷ starting from vRNA residues 216 and 204. After manual fitting of the first several bases into the density, the rest of the connecting loop was built using Rosetta²⁴ and minimized. A model with close fit to the density was chosen for later refinement.

Using the same approach as for the global orientation of the vRNA helices, we find that the fourth and final region of RNA density, located near the RNase H domain of RT, corresponds to remaining portion of tRNA^{Lys}. While both vRNA helices exhibited density consistent with past secondary structure models,

the tRNA density appeared to differ. Instead of revealing expected density for the two independently folded anti-codon and D-stem loops of the tRNA, the RTIC global map showed density consistent with a continuous helix. Also notable was that this helix extends directly from the PBS helix. After re-examining the sequences of the vRNA and tRNA^{Lys}, we noted that it was possible for the vRNA and tRNA^{Lys} to form four extra base pairs, which would extend the PBS from 18 to 22 bp. This extended PBS would be consistent with the continuous helical density in this region. Three out of the four pairs are conserved among subtype-B HIV-1 sequences, suggesting that this structural feature is common. The most variable position would pair with the m1A at position 58 of tRNA^{Lys}. Variability in this position is not unexpected, as a Watson–Crick pair would not be able to form. We generated a second, alternative fold for the remaining portions of the tRNA^{Lys} using mFOLD. This secondary structure, which differed from the previously observed free form secondary by a free energy of less than 1 kcal/mol, is consistent with a long helical structure and accounts for the density observed in the RTIC global map. In addition to being a good fit for the density, the bulges in this model are consistent with the locations of modified nucleotides that would exist in human tRNA^{Lys}. This model also sequesters the anticodon bases of the tRNA, in agreement with chemical mapping data that suggest that these bases are paired^{13–16,53,54}. We note that the apical portion of the extended tRNA helix has very weak EM density and is likely to be dynamic. This dynamic nature of the tRNA is illustrated by the wide variety of final conformations seen in the 3D classes of the RTIC. Models of the extended tRNA helix (residues 2–53) were built individually, clustered, and fit into the density.

The models of the nineteen-base pair extended-PBS helix, vRNA helix 1, vRNA helix 2, and the extended tRNA helix were grafted together, with connecting regions built *de novo* with FARFAR²⁴. Coordinate restraints were applied based on the initial fits to the density for each of these four regions. These penalties were applied for deviations in positions of more than 10 \AA . The best-scoring models were fit into the density in Chimera and a single model was manually selected for further refinement. Regions with the worst agreement with the density, as observed by manual inspection, were subjected to further iterations of FARFAR²⁴ rebuilding and density fitting. The final vRNA–tRNA model was merged with a poly-alanine backbone RT model and refined with one round of Phenix real-space refinement⁴⁹ using secondary structure restraints. Owing to the inclusion of all vRNA and tRNA bases found in our RNA constructs into the model, the model building and refinement procedure may force potentially disordered regions to fit into the 8.0 \AA cryo-EM density map. We stress that the global model presented in the manuscript is meant to aid in interpreting the orientation of the vRNA and tRNA helices with respect to RT and its active sites while showing that the density can encompass most RNA elements. The model should not be used to interpret individual base locations or conformations. For the creation of models for classes 3, 4, and the Mg^{2+} , models of the vRNA H1, vRNA H2, and tRNA were taken from the global model described above. These helical regions were treated as rigid bodies and only the connecting hinge regions (Extended Data Fig. 7c) were rebuilt using the protocol described above. All figures for the RTIC core and global models were prepared in Chimera⁵².

Activity assays. For all activity assays, the RTIC, RT, and vRNA–tRNA were purified as described above.

Time-course assay. RTIC (200 nM) was preincubated for 20 min at 37°C in 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 2.5 mM MgCl_2 . Free vRNA–tRNA (200 nM) and RT (2 μM) were also preincubated for 20 min under the same conditions, but with dCTP in order to fully incorporate the first nucleotide before dTTP incorporation. Incorporation reactions were started by adding a mixture of $\alpha\text{-}^{32}\text{P}$ -dTTP (50 nM), and dTTP (50 μM). Reactions were quenched at a range of times from 1 s to 4 h with the addition of EDTA and SDS loading buffer. The reactions were run on an 4–20% SDS–PAGE gel, dried, and exposed for 18 h on a phosphorimager screen (Molecular Dynamics) and imaged with a Storm 860 (Molecular Dynamics). Bands were quantified using ImageQuant. Intensity was normalized to the highest band intensity for the individual time course assays after background subtraction (set to 1). All time course assays were reliably reproduced and the slow reactions required no special equipment⁸. Plotting and curve fitting was done using IgorPro. For NNRTI experiments, 1 μM nevirapine was added to the pre-reaction incubation mixture of the RTIC.

Relative total incorporation assay. Reactions were performed as described above. RTIC reaction mixtures were quenched at 1 h and the free RT and vRNA–tRNA were quenched at 30 min. Samples were quantified as described above. Incorporation was normalized to the average free RT + vRNA–tRNA band intensity (set to 100%). Relative total incorporation assays were done in triplicate.

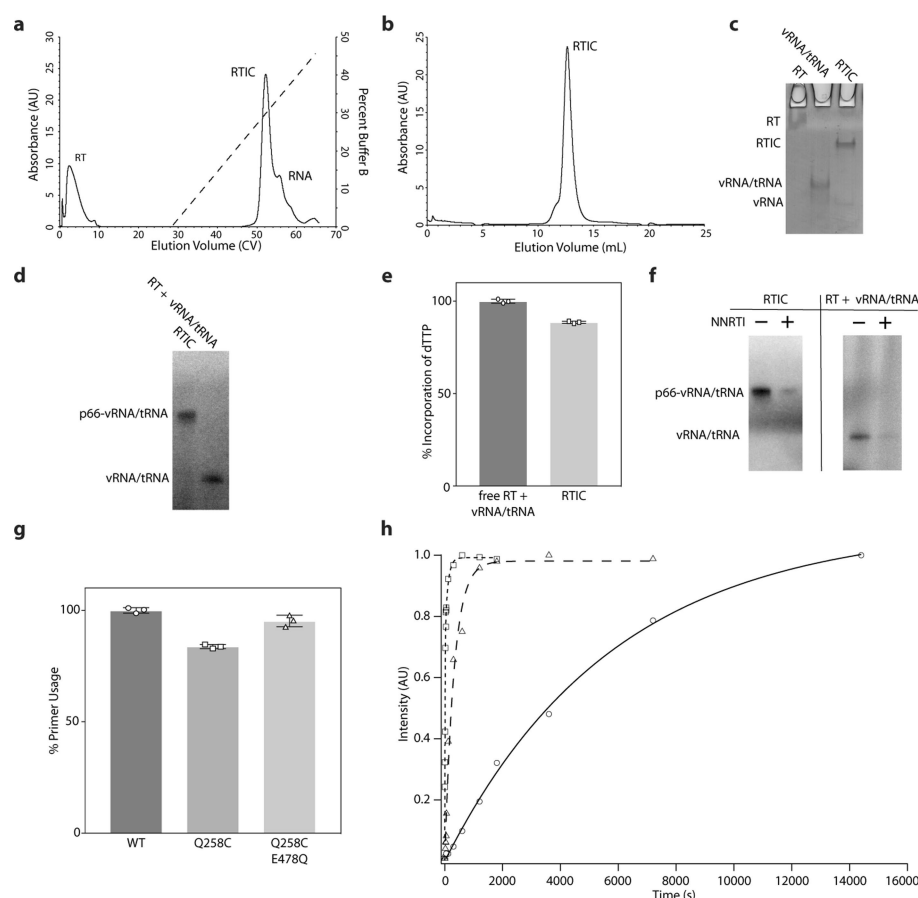
Reverse transcriptase assay. vRNA–tRNA complexes were purified as described above using a tRNA that was labelled on the 5' end with cyanine3 dye. Reactions were pre-incubated at 37°C for 5 min in 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 6 mM MgCl_2 , and 5 mM β -met at a vRNA–tRNA concentration of 200 nM and RT concentration of 3 μM . Reactions were initiated by the addition of a dNTP mixture

that brought the final individual dNTP concentrations to 100 μ M. Reactions were performed in triplicate and quenched at 30 min with EDTA (50 mM). Samples were denatured in a formamide loading buffer, heated for 5 min at 95 °C, and loaded on an 8.5% polyacrylamide gel that was pre-run for 2 h. Samples were run for 3 h at 120 W before imaging with a Typhoon Trio (Amersham Biosciences). Fully extended and unextended primer bands were quantified using ImageQuant. Percent primer extension was calculated and normalized to wild-type RT.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

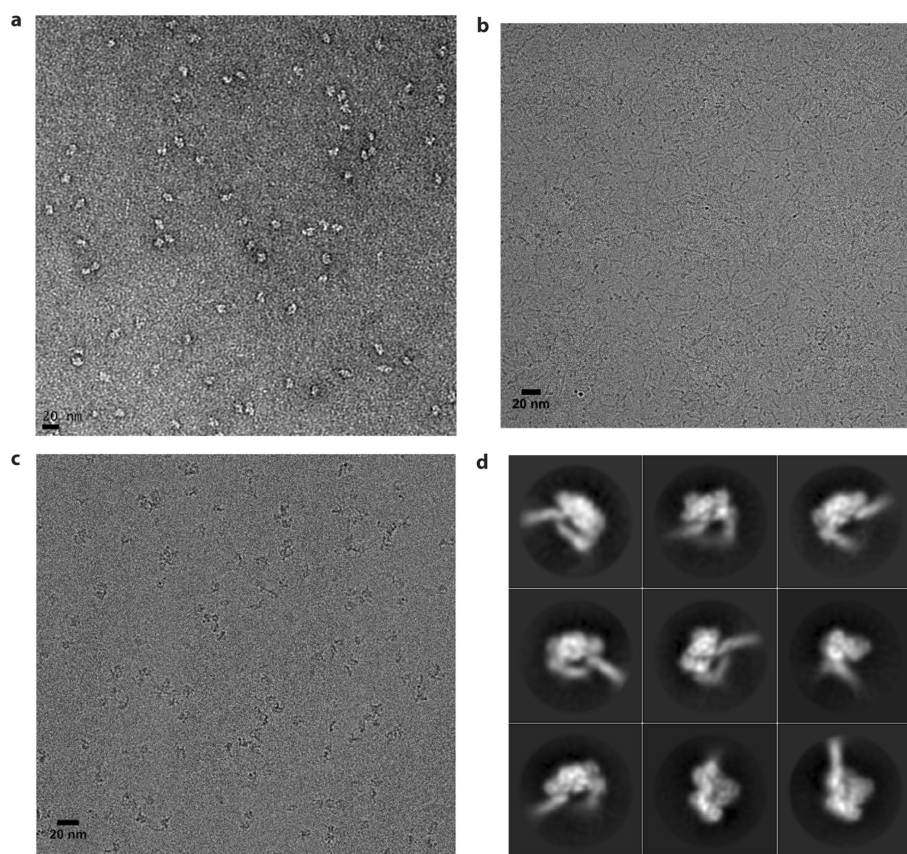
Data availability. Cryo-EM maps of the global RTIC, the core of the RTIC, and the global RTIC with MgCl₂ have been deposited in the Electron Microscopy Data Bank under accession codes EMD-7032, EMD-7031 and EMD-7540. The coordinates of the RTIC core model have been deposited in the Protein Data Bank under accession code 6B19. The global RTIC model is available as Supplementary Data. All other data are available from the corresponding author upon reasonable request.

32. Götte, M. et al. HIV-1 reverse transcriptase-associated RNase H cleaves RNA/RNA in arrested complexes: implications for the mechanism by which RNase H discriminates between RNA/RNA and RNA/DNA. *EMBO J.* **14**, 833–841 (1995).
33. Marshall, R. A., Dorywalska, M. & Puglisi, J. D. Irreversible chemical steps control intersubunit dynamics during translation. *Proc. Natl Acad. Sci. USA* **105**, 15364–15369 (2008).
34. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835 (2008).
35. Johansson, M., Chen, J., Tsai, A., Kornberg, G. & Puglisi, J. D. Sequence-dependent elongation dynamics on macrolide-bound ribosomes. *Cell Rep.* **7**, 1534–1546 (2014).
36. O'Leary, S. E., Petrov, A., Chen, J. & Puglisi, J. D. Dynamic recognition of the mRNA cap by *Saccharomyces cerevisiae* eIF4E. *Structure* **21**, 2197–2207 (2013).
37. Aitken, C. E. & Puglisi, J. D. Following the intersubunit conformation of the ribosome during translation in real time. *Nat. Struct. Mol. Biol.* **17**, 793–800 (2010).
38. Chen, J., Tsai, A., Petrov, A. & Puglisi, J. D. Nonfluorescent quenchers to correlate single-molecule conformational and compositional dynamics. *J. Am. Chem. Soc.* **134**, 5734–5737 (2012).
39. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
40. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
41. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
42. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
43. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
44. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
45. Penczek, P. A., Grassucci, R. A. & Frank, J. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**, 251–270 (1994).
46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
48. Wriggers, W. Conventions and workflows for using Situs. *Acta Crystallogr. D* **68**, 344–351 (2012).
49. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
50. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
51. Lavender, C. A., Gorelick, R. J. & Weeks, K. M. Structure-based alignment and consensus secondary structures for three HIV-related RNA genomes. *PLoS Comput. Biol.* **11**, e1004230 (2015).
52. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
53. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. & Marquet, R. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.* **247**, 236–250 (1995).
54. Isel, C. et al. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA^{3Lys}. *EMBO J.* **15**, 917–924 (1996).



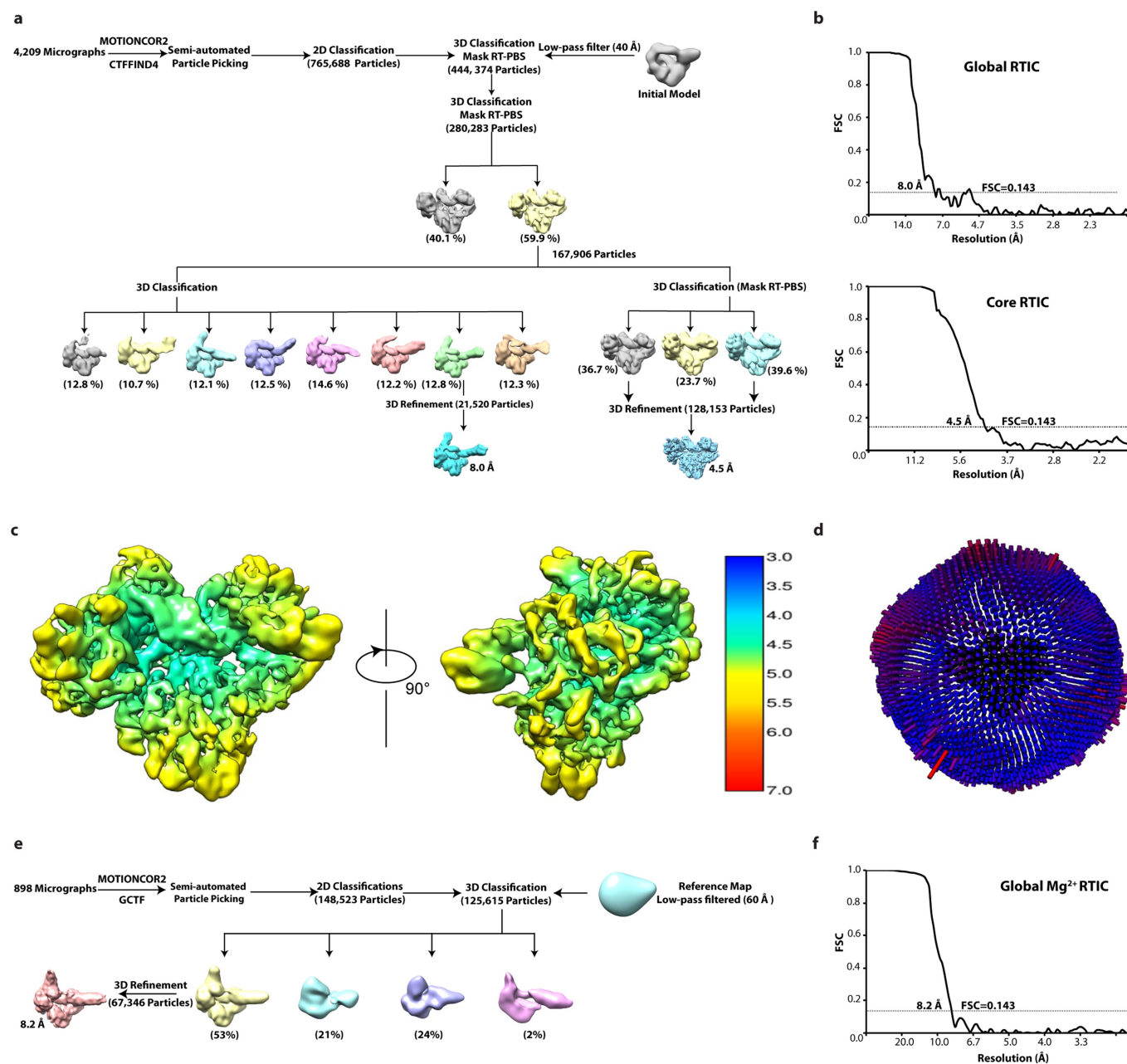
Extended Data Fig. 1 | Purification and activity of RTIC. **a**, Initial anion-exchange purification of the RTIC away from free RT and vRNA-tRNA. This purification was repeated for each sample (>10) used in the manuscript, with only slight variations in the chromatogram. **b**, Polishing step using size-exclusion chromatography purification of the RTIC after anion exchange. This purification was repeated for each sample used in the manuscript (>10), with only slight variations in the chromatogram. **c**, A final 10% native TBE gel on the purified components. RT barely enters the gel under these running conditions. The RTIC runs as a single band, but trace amounts of free vRNA and/or vRNA-tRNA complex are sometimes present. This native gel is a representative result that was repeated independently for all purified RTIC samples used in the paper (>10). **d**, Autoradiograph image illustrating that the RTIC is capable of incorporating an incoming α - 32 P-dTTP nucleotide when extended and purified using dCTP instead of ddCTP. This gel is a representative result that was repeated independently for crosslinked and uncrosslinked samples (>6 independently prepared samples) used in dTTP incorporation assays. **e**, The RTIC incorporates α - 32 P-dTTP at roughly 89% efficiency compared to the free components after reaching a plateau. Values are mean \pm s.d. ($n = 3$ independent experiments) with normalization to total incorporation of free RT + vRNA-tRNA reactions. **f**, Autoradiograph image showing that the incorporation of dTTP is inhibited in the presence of nevirapine (NNRTI). Images have been

adjusted to allow identification of the NNRTI-inhibited band. This gel is a representative result that was repeated independently for crosslinked and uncrosslinked samples (3 samples each). **g**, Relative activities, judged by primer usage, of wild-type, Q258C, and Q258C/E478Q reverse transcriptase mutants used in this study. Values are mean \pm s.d. ($n = 3$ independent experiments) with normalization to the primer usage of wild-type RT. **h**, RTIC (triangles), RTIC with NNRTI (circles) or vRNA-tRNA + excess RT (squares) reactions were initiated by addition of α - 32 P-dTTP and quenched at different time points. Data were fit using the relationship for the free vRNA-tRNA + RT reaction: $\text{Intensity} = A(1 - e^{-k_{\text{pol}}t}) + B(1 - e^{-k_{\text{slow}}t})$. Data were fit using the relationship for the RTIC (with or without NNRTI) reaction: $\text{Intensity} = B(1 - e^{-k_{\text{slow}}t})$ where A and B represent the amplitude of the fast and slow processes, respectively, k_{pol} is the apparent extension rate constant, and k_{slow} is the rate of the slow process. The second relationship was used for the RTIC data, as the slow process appears to dominate incorporation when the vRNA-tRNA substrate is crosslinked to RT. The best fits were obtained with: $A = 0.7166$ AU, $k_{\text{pol}} = 0.1078 \text{ s}^{-1}$, $B = 0.2754$, $k_{\text{slow}} = 0.01002 \text{ s}^{-1}$ for the vRNA-tRNA + excess RT; $B = 0.9808$, $k_{\text{slow}} = 0.003140 \text{ s}^{-1}$ for the RTIC; and $B = 1.095$, $k_{\text{slow}} = 0.0001714 \text{ s}^{-1}$ for the RTIC with NNRTI. k_{slow} is about 3.19 times slower for crosslinked RTIC than for un-crosslinked components. Assays were independently repeated three times to ensure reproducibility.



Extended Data Fig. 2 | Representative negative-stain EM images, cryo-EM images, and 2D averages of the RTIC. a, Representative negative-stain EM image of HIV RTIC reveals a mono-disperse sample that is free of aggregates. Approximately a dozen images were taken of each sample before cryo-EM grid preparation to ensure sample quality. **b,** Cryo-EM image of RTIC without β -OG. The long chains correspond to RNA from the complex with very few particles resembling the protein.

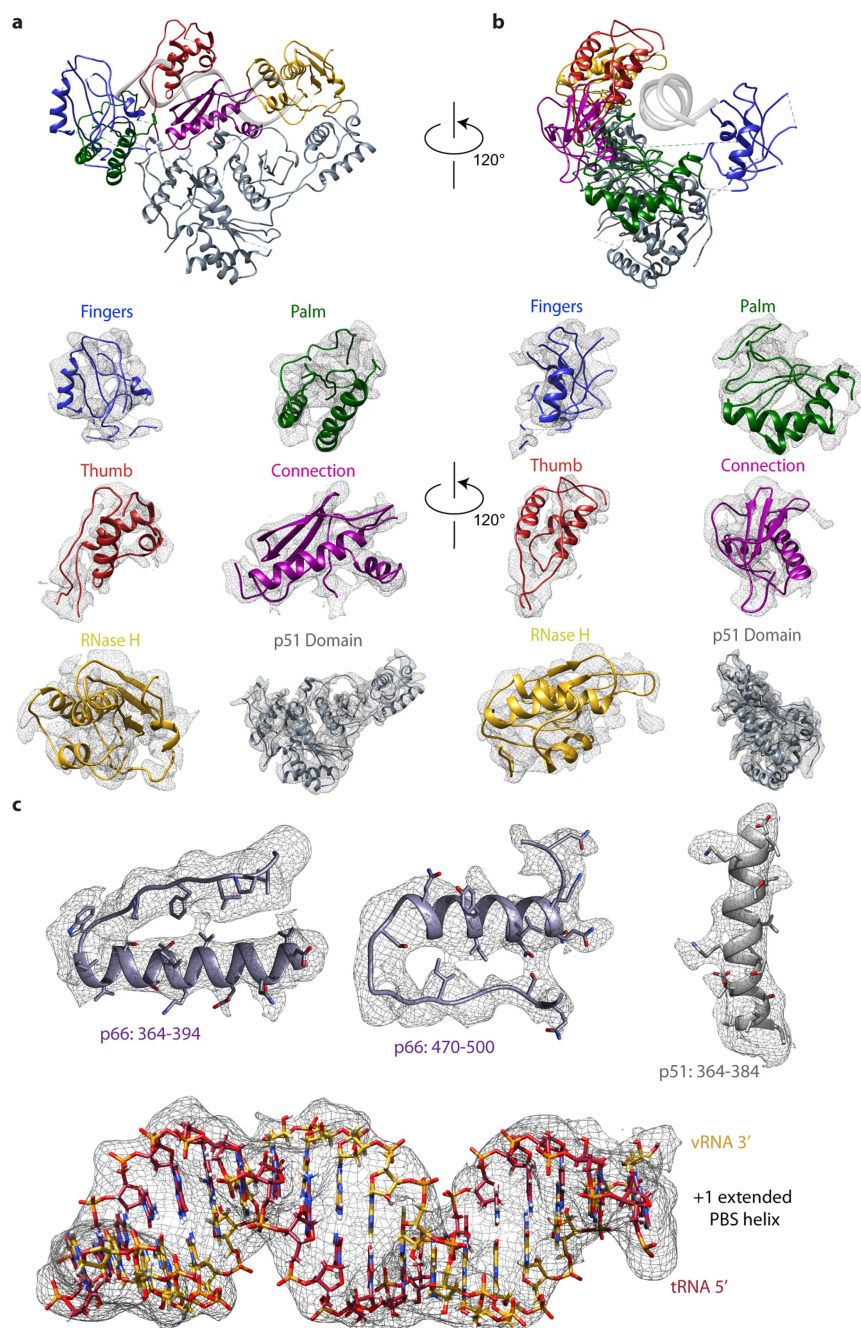
Results are reproducible in the absence of β -OG (>10 samples tested). **c,** Cryo-EM image of RTIC with β -OG. Single particles corresponding to the complex appear similar to the negative-stain visualization. All 5,107 images used in both cryo-EM datasets have a similar appearance with slight differences in particle density. **d,** Representative 2D averages of RTIC complex from the cryo-EM data collected with β -OG. Both datasets exhibit very similar 2D classes.



Extended Data Fig. 3 | Data processing workflow for RTIC complex.

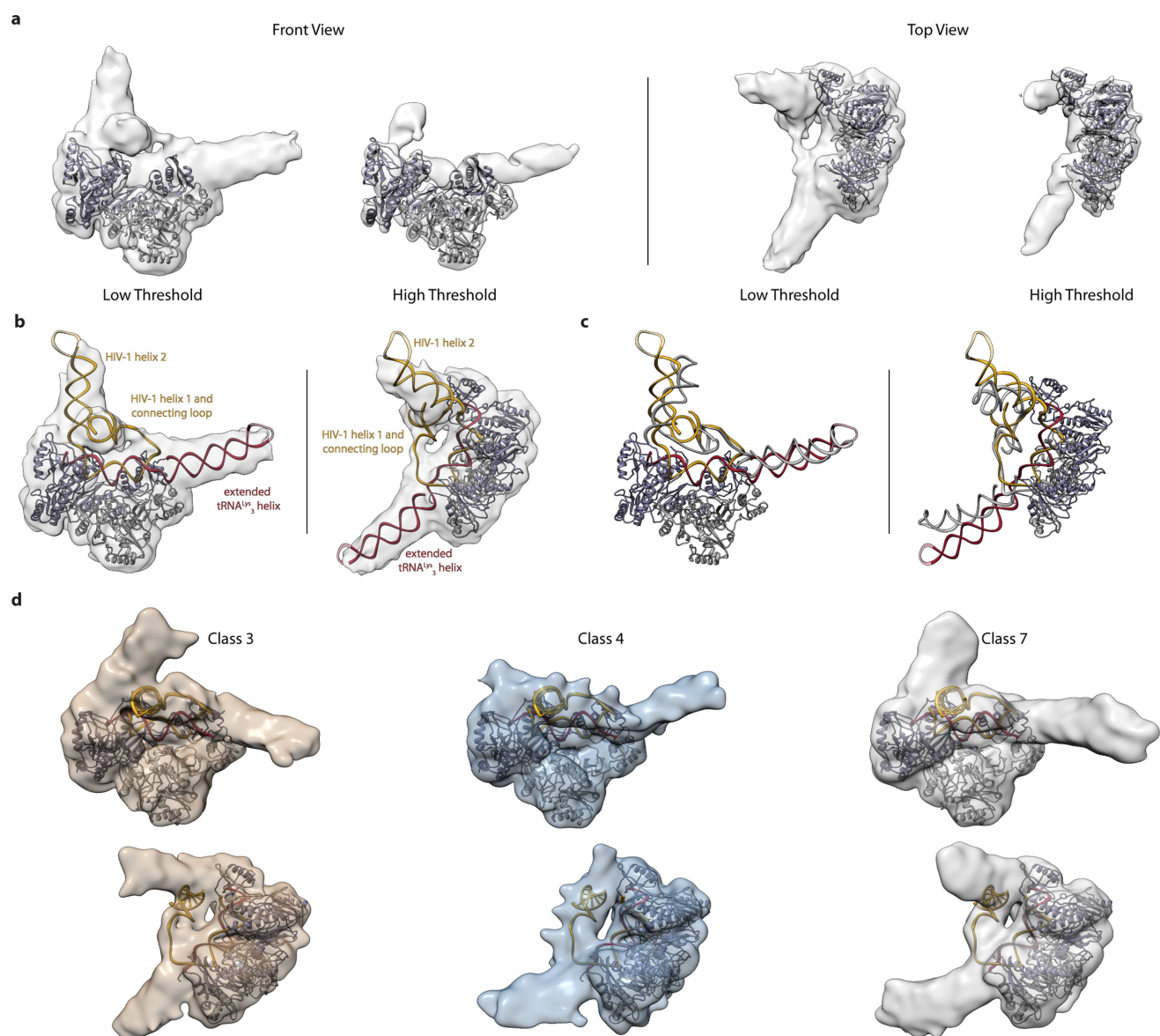
a, Data processing workflow for the 8.0 Å global and 4.5 Å core maps. **b**, Gold standard FSC curve of RTIC core and global maps. **c**, The final 4.5 Å map is coloured according to local resolution estimated by Relion.

d, Angular distribution of particle projections. The length of each projection direction is proportional to the number of assigned particles. **e**, Data processing workflow for the 8.2 Å global Mg^{2+} map. **f**, Gold standard FSC curve of RTIC Mg^{2+} global map.



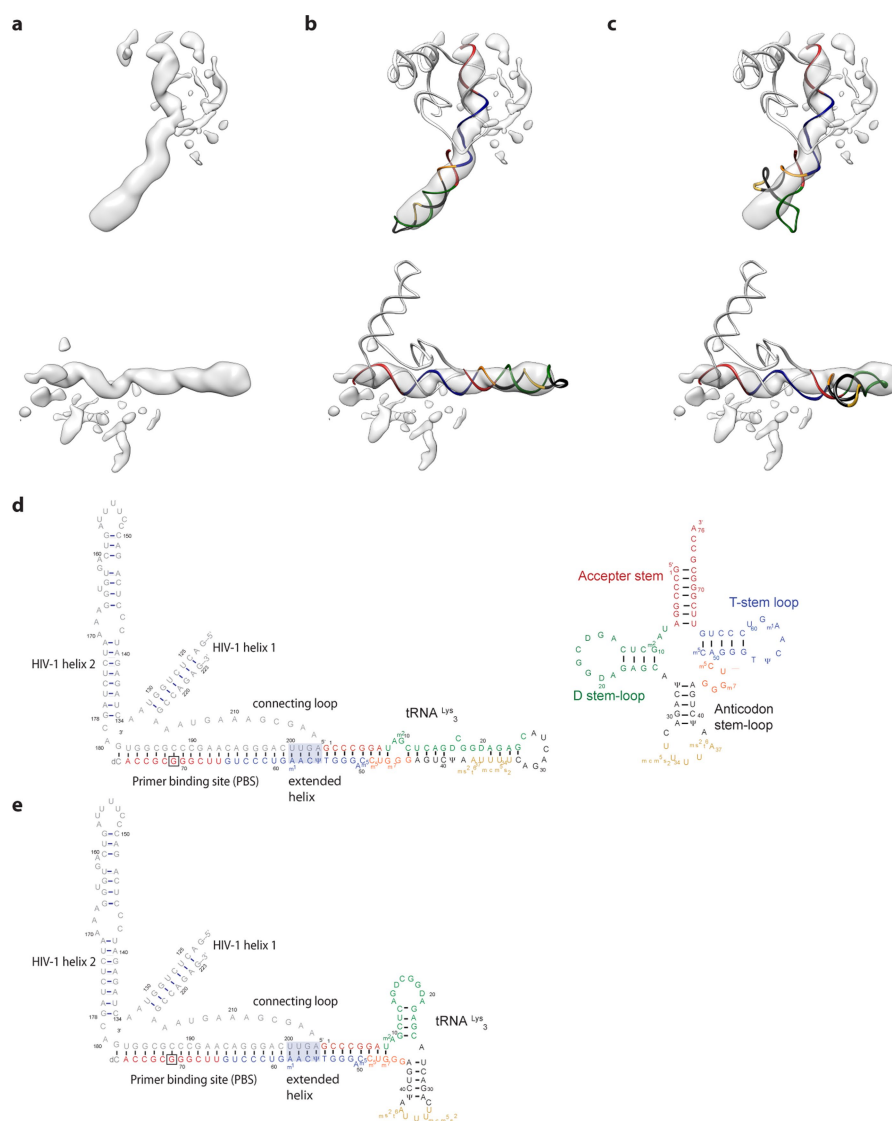
Extended Data Fig. 4 | Quality of the cryo-EM density for the core RTIC map. **a**, View of HIV-1 RT from the front. The subdomains of RT are coloured. Underneath the main RTIC view, each subdomain of RT, plus the p51 subunit, is shown fit into the 4.5 Å map. **b**, View of HIV-1 RT from the polymerase active site side. The subdomains of RT are coloured. Underneath the main RTIC view, each subdomain of RT, plus the p51 subunit, is shown fit into the 4.5 Å map. In **a**, **b**, regions of protein, namely loops and linkers, that lacked sufficient density were removed after

comparison with previously published structures of RT. These regions are indicated by dotted lines and are most commonly found in the finger and palm subdomains. **c**, Representative regions of 4.5 Å map fitted with protein secondary structure that display densities for side chains. A view of the PBS helix fit into the 4.5 Å map is also shown; phosphates of the RNA backbone are partially resolved. Regions are coloured with respect to the main text models.



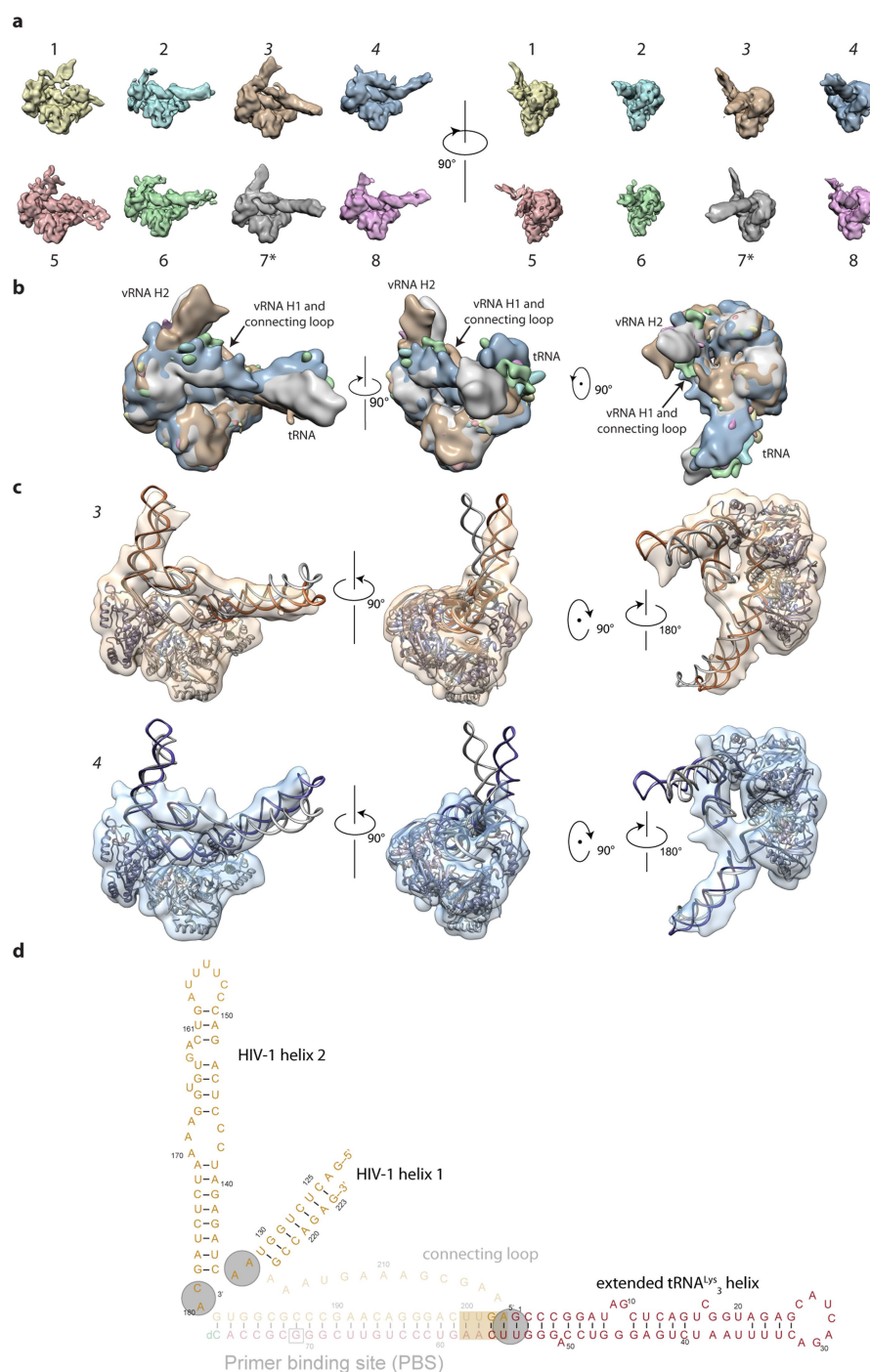
Extended Data Fig. 5 | Mg^{2+} global map views and structure comparison. **a**, Side and top views of the 8.2 Å global map at different density thresholds. The orientation of the peripheral vRNA and tRNA elements is within the variability seen among the different RTIC conformers. **b**, A model of the RTIC built into the Mg^{2+} density using the main text global RTIC model. vRNA and tRNA helices were treated as rigid bodies derived from main text model (see Extended Data Fig. 6 and

Methods). **c**, Comparison of the global RTIC model RNA (grey) with the Mg^{2+} model RNA (coloured). All three regions of RNA structure (H1, H2, and tRNA) differ in the Mg^{2+} model, but are adequately described by rigid body movements of the RNA helical elements taken from the global RTIC model. Both H1 and H2 represent a substantial structural barrier to initiation. **d**, Partial accommodation of H1 into high monovalent salt classes 3, 4 and 7.



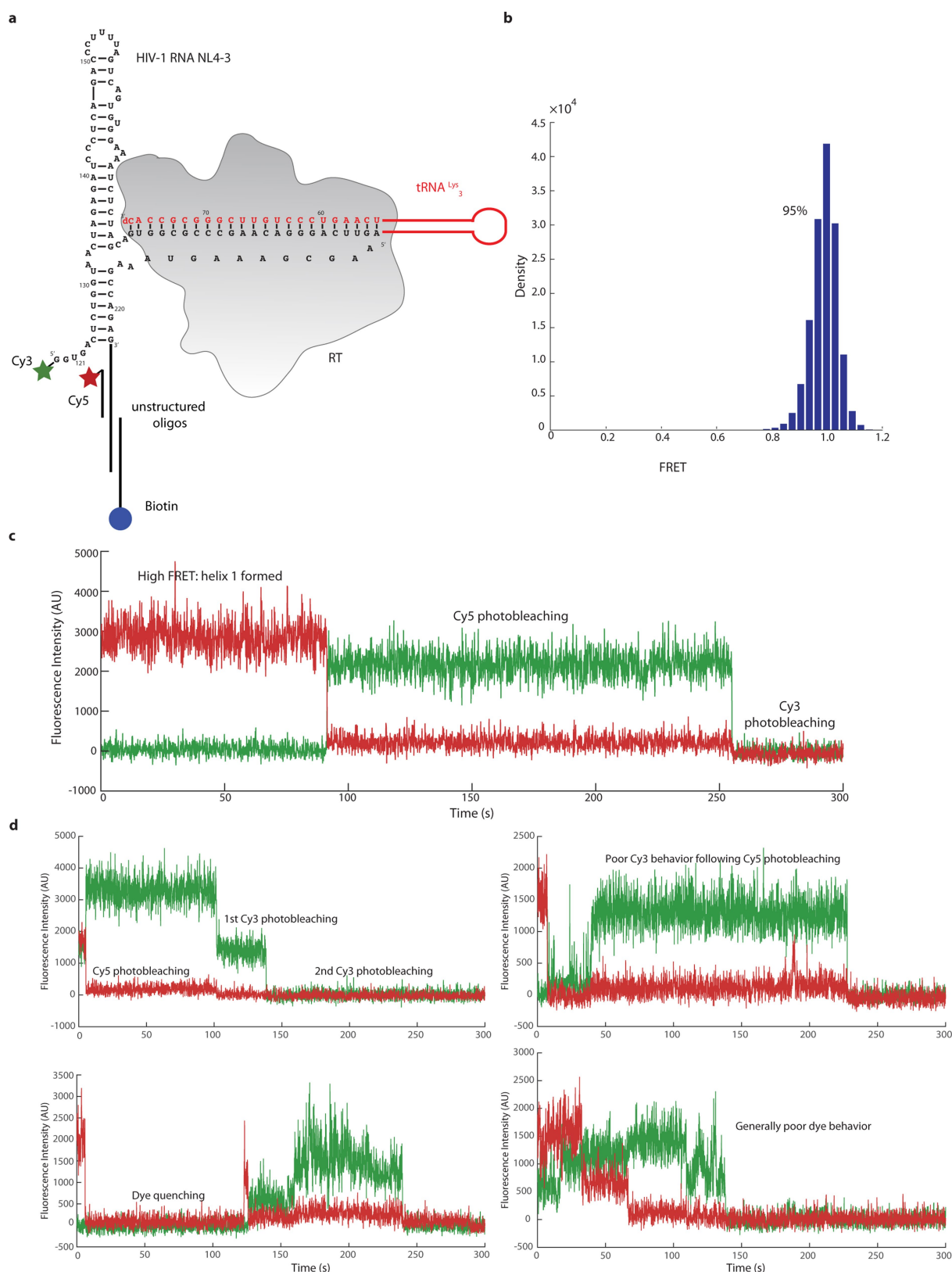
Extended Data Fig. 6 | Low-resolution tRNA density and fold comparison. **a**, Top and side views of the elongated helical tRNA density observed in the low-resolution global map of the RTIC. **b**, Top and side views of the vRNA–tRNA model generated using the hypothesized elongated tRNA helical fold. The tRNA model fits the long helical density well. Corresponding secondary structure is in **d**. **c**, Top and side views of the vRNA–tRNA model generated using previously hypothesized tRNA secondary structures that have the anticodon and D-stem loops

independently folded. Corresponding secondary structure is in **e**. **d**, Secondary structure depiction of the new vRNA–tRNA and canonical clover-leaf fold of the tRNA. The different domains are coloured and correspond with the models in panels **b** and **c**. **e**, Secondary-structure depiction of the old vRNA–tRNA fold with independent anticodon and D-stem loops. The domains are coloured and correspond with the model in **c** and clover-leaf fold of the tRNA in **d**.



Extended Data Fig. 7 | Peripheral RNA heterogeneity of the RTIC conformers. **a**, Tiled views of eight conformations emerging from 3D classification of RTIC. Each class is numbered and class 7 was used for the global RTIC reconstruction. **b**, Superposition of the eight classes from **a**. The main areas of RNA heterogeneity are focused on the orientations of vRNA H2, H1 and the connection loop, and the tRNA. With no stabilizing protein contacts, vRNA H2, H1, and the tRNA sample a wide range of conformations, limiting the resolution of the global map. **c**, Additional RTIC models built into classes 3 (tan) and 4 (blue). The models for the

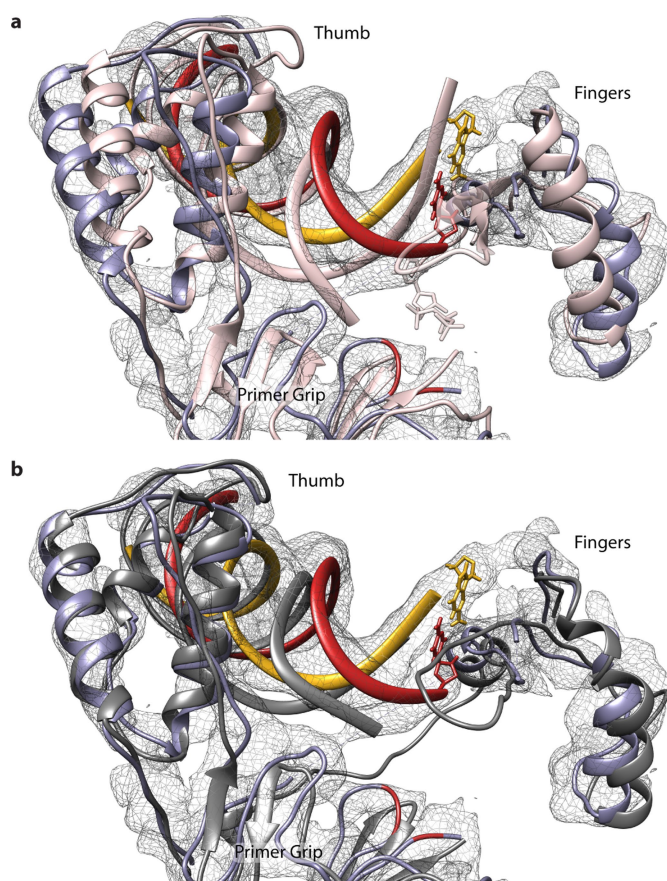
tRNA, vRNA H1, and vRNA H2 were all derived from the global RTIC model and treated as rigid bodies for model building. The connecting loop was not built in these models as the density for this region was not clear in these maps, though there is reasonable density to model a loop near H1. Junctions between the helices serve as hinges that allow movement of the independent domains. The main text global RTIC model (grey) is included as a comparison. **d**, The vRNA and tRNA helices treated as rigid bodies for modelling are shown in bold. Hinge points for each helix are highlighted with grey circles and serve as points of flexibility for the RTIC.



Extended Data Fig. 8 | Single-molecule experimentation and analysis.

a, Secondary structure depiction of the vRNA-tRNA construct used for single-molecule experiments. The labelling scheme is shown, with the Cy3 dye located on the 5' end of the vRNA helix 1 and Cy5 dye located on an oligonucleotide positioned near the 5' end of helix 1. The vRNA-tRNA complex was crosslinked to RT for the experiments. **b**, Ninety-five per cent of the RTIC complexes are in the high FRET, helix 1 formation, state

(480 traces analysed, see Methods). **c**, Example trace of the ones used for final FRET analysis. The high FRET state of the RTIC complex, which is attributed to helix 1 formation. Photobleaching events for both Cy5 and Cy3 are indicated. **d**, Examples of traces removed from final FRET analysis. Traces exhibit the presence of multiple molecules (multiple single-dye photobleaching events) or poor dye behaviour (blinking and quenching).



Extended Data Fig. 9 | Comparison with NNRTI bound and active RT–nucleic acid complexes in the cryo-EM map. All alignments between structures and the RTIC were done using the p51 subunit. **a**, Comparison of an active conformation RT–nucleic acid structure (pink, 1RTD) with the RTIC core (RT, purple; tRNA primer, red; vRNA template, yellow). The EM map overlay shows the poor fit of the 1RTD model in the fingers, thumb, and primer grip of RT. Deviations of the nucleic acid primer and template of 1RTD away from the RTIC density are also apparent. **b**, Comparison of an NNRTI-bound RT–nucleic acid structure (dark grey, 3V81) with the RTIC core. The EM map overlay shows the closer fit of the fingers and primer grip regions of RT in the 3V81 model. The thumb region also overlays well, but with slight deviations. Most noticeably, the nucleic acid primer/template in the 3V81 model deviates, although not as dramatically as in 1RTD, from the RTIC core EM density.

Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics

	RTIC Core (EMDB-7031) (PDB-6B19)	RTIC Global (EMDB-7032) (Supplementary Data file for model)	RTIC Global w/MgCl ₂ (EMDB-7540)
Data collection and processing			
Magnification (calibrated)	50,000	50,000	38,880
Voltage (kV)	300	300	200
Electron exposure (e-/Å ²)	60 and 85	60 and 85	75
Defocus range (μm)	-1.3 to -2.5	-1.3 to -2.5	-2.0 to -3.0
Pixel size (Å)	1.0	1.0	1.286
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	765,688	765,688	148,523
Final particle images (no.)	128,153	21,520	67,346
Map resolution (Å)	4.5	8.0	8.2
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used (PDB code)	3V81	3V81	
Map sharpening <i>B</i> factor (Å ²)	-250	-200	-200
Model composition			
Non-hydrogen atoms	5,299	8,545	
Protein residues	909	962	
RNA nucleotides	38	178	
R.m.s. deviations			
Bond lengths (Å)	0.023	0.003	
Bond angles (°)	1.603	0.83	
Validation			
MolProbity score	2.41	1.92	
Clashscore	15.00	14.25	
Poor rotamers (%)	N/A	N/A	
Ramachandran plot			
Favored (%)	80.46	96.15	
Allowed (%)	19.42	3.85	
Disallowed (%)	0.12	0.00	

Structure of the alternative complex III in a supercomplex with cytochrome oxidase

Chang Sun^{1,2,9}, Samir Benlekbi^{3,9}, Padmaja Venkatakrishnan^{1,8,9}, Yuhang Wang^{1,2,4}, Sangjin Hong¹, Jonathan Hosler⁵, Emad Tajkhorshid^{1,2,4,10*}, John L. Rubinstein^{3,6,7,10*} & Robert B. Gennis^{1,10*}

Alternative complex III (ACIII) is a key component of the respiratory and/or photosynthetic electron transport chains of many bacteria^{1–3}. Like complex III (also known as the *bc*₁ complex), ACIII catalyses the oxidation of membrane-bound quinol and the reduction of cytochrome *c* or an equivalent electron carrier. However, the two complexes have no structural similarity^{4–7}. Although ACIII has eluded structural characterization, several of its subunits are known to be homologous to members of the complex iron–sulfur molybdoenzyme (CISM) superfamily⁸, including the proton pump polysulfide reductase^{9,10}. We isolated the ACIII from *Flavobacterium johnsoniae* with native lipids using styrene maleic acid copolymer^{11–14}, both as an independent enzyme and as a functional 1:1 supercomplex with an *aa*₃-type cytochrome *c* oxidase (cyt *aa*₃). We determined the structure of ACIII to 3.4 Å resolution by cryo-electron microscopy and constructed an atomic model for its six subunits. The structure, which contains a [3Fe–4S] cluster, a [4Fe–4S] cluster and six haem *c* units, shows that ACIII uses known elements from other electron transport complexes arranged in a previously unknown manner. Modelling of the cyt *aa*₃ component of the supercomplex revealed that it is structurally modified to facilitate association with ACIII, illustrating the importance of the supercomplex in this electron transport chain. The structure also resolves two of the subunits of ACIII that are anchored to the lipid bilayer with N-terminal triacylated cysteine residues, an important post-translational modification found in numerous prokaryotic membrane proteins that has not previously been observed structurally in a lipid bilayer.

The ACIII–cyt *aa*₃ supercomplex from *F. johnsoniae* membranes was solubilized, purified and biochemically characterized using styrene maleic acid (SMA) copolymer nanodiscs without traditional detergents (Supplementary Discussion, Extended Data Figs. 1–3). The supercomplex catalyses the two-electron oxidation of menaquinol (or ubiquinol) and the four-electron reduction of oxygen to water with a turnover number of around 21 electrons per second without the addition of exogenous cyt *c* (Supplementary Information, Extended Data Fig. 3), indicating a functional electron transfer chain within the supercomplex. The addition of exogenous cyt *c* did not increase the rate of electron transfer. The structure of the ACIII–cyt *aa*₃ supercomplex in SMA nanodiscs was determined by cryo-electron microscopy (cryo-EM) (Fig. 1, Extended Data Fig. 4). The supercomplex has a mass of 464 kDa (Supplementary Discussion), a transmembrane cross-section of approximately 9 nm × 13 nm (Extended Data Fig. 5), and contains 48 transmembrane α -helices. To our knowledge, the ACIII–cyt *aa*₃ supercomplex is the largest protein complex reported to be contained within an SMA copolymer nanodisc. The SMA copolymer and lipids contribute only a thin layer of density around the supercomplex (Fig. 1a, b), which is not circular but follows the contours of the protein.

Whether this is a general feature of SMA-solubilized proteins or is due to the large size of the ACIII–cyt *aa*₃ supercomplex is not known, and will be clarified when more structures are determined using this approach. The number of loosely bound, unresolved lipid molecules is not known, nor is it known whether they are sufficient in number to form a true bilayer surrounding the protein. The SMA–supercomplex nanodiscs retain native lipids, are more stable and have 30% higher specific activity than the supercomplex isolated with detergents (for example, dodecylmaltoside) (Supplementary Discussion, Extended Data Fig. 3). Because traditional detergents are avoided in generating SMA nanodiscs, the preparative protocol is more rapid and simpler than making nanodiscs using membrane scaffold proteins.

Although the properties of the SMA nanodiscs are less well characterized than nanodiscs made with membrane scaffold proteins^{13,15}, our work demonstrates the utility of SMA nanodiscs for high-resolution structural studies of membrane proteins.

The resolution of the cryo-EM density map enabled construction of an atomic model for more than 90% of the sequences predicted from the ACIII gene cluster (Supplementary Discussion), including subunits ActA, ActB, ActC, ActD, ActE and ActF (Fig. 2, Extended Data Fig. 5, Extended Data Table 1). Sequence analysis shows that ACIII contains a unique combination of known modules from other respiratory complexes³ (Supplementary Discussion). The ACIII structure confirms this prediction and shows the structure responsible for catalysing the quinol:cyt *c* oxidoreductase activity. The ACIII structure can be divided into three parts: a core assembly of ActC and ActB that oxidizes quinol; a haem *c* assembly consisting of ActA and ActE that directs electrons from ActB to the terminal electron acceptor; and auxiliary transmembrane subunits ActD and ActF with unknown functions. With some key differences (Extended Data Fig. 5), the overall architecture of ActB and ActC resembles the complex consisting of the PsrA, PsrB and PsrC

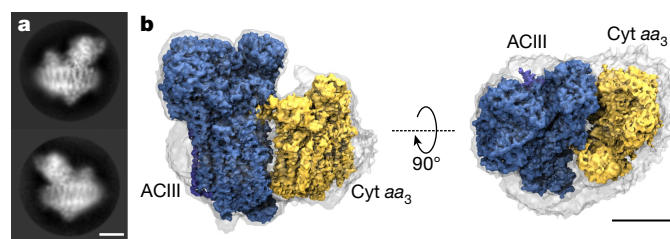


Fig. 1 | Cryo-EM of the ACIII–cyt *aa*₃ supercomplex in SMA nanodiscs. **a**, Two representative 2D class average images of the ACIII–cyt *aa*₃ supercomplex in a nanodisc. Calculation of 2D class averages was not repeated. **b**, Side (left) and top (right) views of the ACIII–cyt *aa*₃ supercomplex cryo-EM map. The transparent surface indicates the boundary of the nanodisc. Scale bars, 50 Å.

¹Department of Biochemistry, University of Illinois, Urbana, IL, USA. ²NIH Center for Macromolecular Modeling and Bioinformatics, Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, IL, USA. ³Molecular Medicine Program, The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada. ⁴Center for Biophysics and Quantitative Biology, University of Illinois, Urbana, IL, USA. ⁵Department of Biochemistry, University of Mississippi Medical Center, Jackson, MS, USA. ⁶Department of Medical Biophysics, The University of Toronto, Toronto, Ontario, Canada. ⁷Department of Biochemistry, The University of Toronto, Toronto, Ontario, Canada. ⁸Present address: Department of Microbiology and Molecular Genetics, University of California, Davis, Davis, CA, USA. ⁹These authors contributed equally: C. Sun, S. Benlekbi, P. Venkatakrishnan. ¹⁰These authors jointly supervised this work: E. Tajkhorshid, J. L. Rubinstein, R. B. Gennis. *e-mail: emad@life.illinois.edu; john.rubinstein@utoronto.ca; r-gennis@illinois.edu

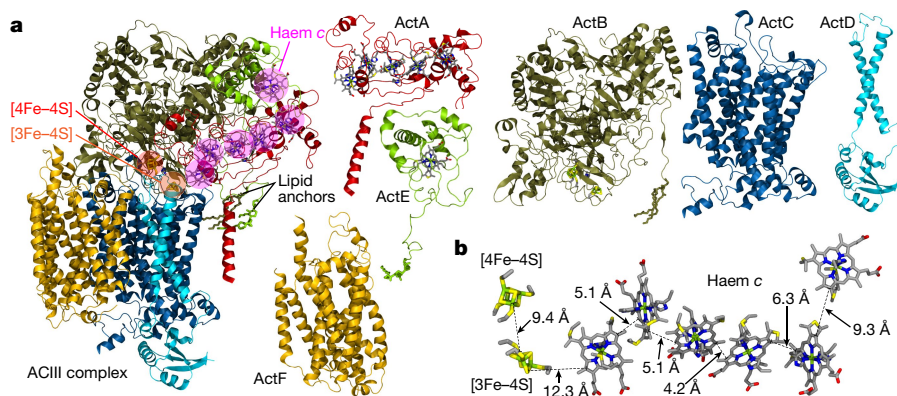


Fig. 2 | Atomic model of ACIII. **a**, The overall structure is shown on the left, along with separate views of individual subunits. Covalently bonded cysteines are shown along with the core iron–sulfur clusters (circled in orange). Both covalently linked cysteines and axial-coordinating residues are shown for haems (circled in pink). **b**, The edge-to-edge distances between cofactors.

subunits of polysulfide reductase from *Thermus thermophilus* (PsrABC)¹⁶, a member of the CISM superfamily (Supplementary Discussion). Like PsrC, ActC contains no cofactors, but it does contain the proposed site for the oxidation of menaquinol. Residues at the menaquinol-binding site identified in PsrC¹⁶ are not conserved in ActC¹⁷. Although menaquinone is not observed in the cryo-EM map, we propose that the ActC residues His133 and Asp164 form the menaquinol-binding site in ActC near the interface with ActB (Extended Data Fig. 6). These two residues are conserved in ActC sequences and there is a crevice between transmembrane helices 3 (TM3) and 4 (TM4) of ActC that would provide access to the substrate in the membrane bilayer.

The N-terminal portion of ActB is homologous to the PsrA subunit of polysulfide reductase, which contains the molybdenum cofactor, but the molybdenum cofactor is absent in ActB⁴. The C-terminal domain of ActB is homologous to PsrB, and both ActB and PsrB contain iron–sulfur clusters. Like PsrB, ActB from *F. johnsoniae* is expected to contain four iron–sulfur clusters, but only two are observed in the cryo-EM map (Extended Data Fig. 7). There is one [3Fe–4S] cluster near the interface with ActC, about 10 Å from the proposed site of menaquinol oxidation, and one [4Fe–4S] cluster about 9 Å further away. There are two additional cysteine clusters present in the structure of ActB, but the cryo-EM map does not show iron–sulfur clusters at these locations. Instead, we observe disulfide bonds (Cys965–Cys938 and Cys971–Cys769) within these two cysteine clusters in ActB. The substitution of proposed [4Fe–4S] clusters by disulfide bonds may be a genuine aspect of the structure or may result from oxidation that occurred during sample preparation. However, if these two ‘missing’ [4Fe–4S] clusters were present, they would form a dead-end for electron transfer from the [3Fe–4S] cluster of ActB, suggesting that their absence from the structure is not an artefact.

The [3Fe–4S] cluster in ActB is the most probable initial oxidant of menaquinol bound to ActC, and is 12.3 Å from the nearest haem *c* in ActA. The five haems *c* in ActA plus the single haem *c* in ActE form a probable electron transfer wire from the [3Fe–4S] cluster in ActB, with the largest edge-to-edge distance of 9.2 Å between adjacent haems (Fig. 2b). The [4Fe–4S] cluster in ActB appears to be off-pathway and its function remains to be determined.

In all *Flavobacteria*, including *F. johnsoniae*, ActA is predicted to have a monohaem domain at the N terminus in addition to the pentahaem domain at the C terminus (Supplementary Discussion). Mass spectrometry analysis shows that the N-terminal monohaem domain is present in the preparation (Extended Data Fig. 1), but no density can be assigned to this entire domain. The inability to resolve the monohaem domain may result from flexibility of the domain. Full-atom molecular dynamics simulations were performed for the entire structure of ACIII embedded in a phospholipid bilayer to determine the stability and dynamics of the structure (Extended Data Fig. 8). Notably, the pentahaem domain of ActA had the largest root-mean-square deviation (r.m.s.d.), which arises mainly from the transmembrane α -helix connected to the missing monohaem domain; this is consistent with the

monohaem domain being unobservable owing to a variable position in the complex. Although ActE also had a substantial r.m.s.d., it did not appear to correlate with disorder in the cryo-EM map.

ActD and ActF are transmembrane subunits without bound cofactors, and both interact with ActC. It has not been established whether ACIII generates a proton motive force coupled to electron transport¹⁸. The absence of redox centres in ActC, ActD and ActF suggests that if ACIII contributes to the transmembrane proton gradient, it does not use the bifurcation-type Q-cycle mechanism of canonical complex III¹⁹, but instead functions as a true proton pump with a mechanism that resembles that of complex I²⁰. ActD has two transmembrane α -helices that cross within the membrane and are adjacent to ActC. Both N and C termini are within the cytoplasm and combine to form a single globular domain that rests on the cytoplasmic surface of ActC. The ten transmembrane α -helices of ActF form a pseudo two-fold axis of symmetry with the ten transmembrane α -helices of ActC (Extended Data Fig. 5), despite the fact that ActF has less than 20% sequence identity with ActC. If ACIII is a proton pump, it is likely that conserved polar residues within the bilayer will have important roles.

The structure of ACIII reveals eleven ordered phospholipid molecules as well as triacylated cysteine residues at the N termini of ActB (Fig. 3a) and ActE (Extended Data Fig. 7). The anchoring of bacterial membrane proteins by an N-terminal triacylated cysteine is a well-characterized phenomenon²¹; however, to our knowledge, this is the first time the structure of a triacylated cysteine residue has been determined in the context of a membrane protein. Both lipid anchors are tilted with respect to the plane of the lipid bilayer (Fig. 2a), restricting the ability of other lipids to pack around them. This feature could alter the mechanical properties of the adjacent portion of the membrane bilayer, and also guide conformational changes in the ACIII protein. Notably, the two N-terminal lipid anchors are adjacent to each other in the membrane. These lipid anchors probably help ACIII to assemble and keep the monohaem ActE bound to the complex. The eleven lipids that are resolved adjacent to the transmembrane α -helices accommodate the rugged protein surface of the complex (Fig. 3b, Extended Data Fig. 7). The head groups of the lipids could not be identified and were all modelled as phosphatidylethanolamine. There are two ‘hot spots’ for resolved lipids: the cytoplasmic interface between ActC and ActF; and the vicinity of the triacylated cysteine of ActB, which is near the proposed entry point for menaquinol into the complex. All eleven of the resolved lipids remained bound to the protein throughout 250 ns of molecular dynamics simulation (Extended Data Fig. 8), supporting the ability of SMA nanodiscs to preserve some native lipid–protein interactions and suggesting a functional role for the lipids. A large number of annular lipids, including those modelled in the structure, were observed to associate with the protein from the *in silico* bilayer.

Frequently, the subunits encoding ACIII are within an operon that includes subunits of an associated complex IV³ (cyt *aa*₃ or cyt *caa*₃). We find that the sequences of subunit III from complex IVs that are associated with ACIII have unusual features that distinguish them from the canonical subunit III (Supplementary Discussion).

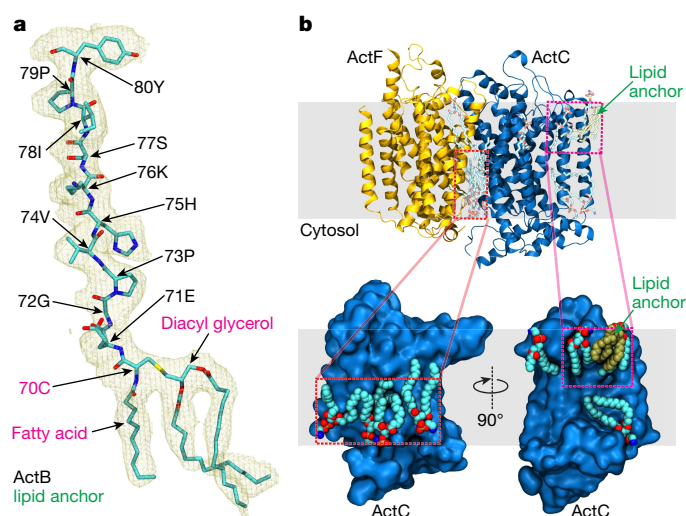


Fig. 3 | Lipids in the structure of ACIII. **a**, Triacylated cysteine at the N terminus of ActB. The triacylated cysteine and its downstream ten amino acid residues are shown in the context of the experimental density map. **b**, Other resolved lipids near the transmembrane α -helices. Four lipid molecules are resolved at the cytoplasmic interface between ActC and ActF. Alongside, two lipid molecules are clustered near the triacylated cysteine from ActB, directly above the proposed quinone entry pathway.

Whereas subunit III of complex IV generally contains seven transmembrane α -helices, those that are associated with ACIII lack TM1 and TM2 (Fig. 4a). Although only parts of subunit III of *cyt aa₃* are resolved to better than 4 Å, the density for *cyt aa₃* has sufficient resolution to identify five α -helices from the structure. A homology model of subunit III from *F. johnsoniae* *cyt aa₃* was built on the basis of the structure of TM3 to TM7 of subunit III from *Rhodobacter sphaeroides* *cyt aa₃*, and fit into the ACIII–*cyt aa₃* supercomplex density map (Extended Data Fig. 9) with high fidelity. The deletion of the first two transmembrane α -helices in subunit III of *cyt aa₃* appears to be a necessary adaptation to enable formation of the supercomplex with ACIII. It is notable that the same two helices in subunit III are also absent in the *cyt aa₃* obligatory *cyt bcc*–*cyt aa₃* supercomplex found in *Actinobacteria* (for example, *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*)²².

The sequence analysis also reveals that the loop between TM5 and TM6 of subunit III in the *cyt aa₃* that is part of the supercomplex is much longer in *F. johnsoniae* (and all *Flavobacteria*) than in other organisms. Typically, this loop contains eight residues, but in *F. johnsoniae* it contains 121 residues (Fig. 4a). Part of this long loop fits in a groove between ActB and ActD of ACIII on the periplasmic side of the membrane (Extended Data Fig. 9). The structural model reveals a π -cation interaction between Trp188 of subunit III and Arg868 of ActB (Fig. 4b), both of which are conserved among organisms containing subunit III with a long loop between TM5 and TM6 (Extended Data Fig. 9). This specific and strong interaction stabilizes the ACIII–*cyt aa₃* supercomplex and appears to be a second adaptation that enables the formation of a supercomplex with ACIII. The contact between the periplasmic loop of subunit III of *cyt aa₃* and ACIII is the only observed direct contact between the two complexes. The five well-resolved transmembrane α -helices of subunit III of *cyt aa₃* are angled away from ACIII with only the tip of TM6 of subunit III touching ActF, forming a wedge-like space between the membrane domains of ACIII and *cyt aa₃*. The decrease of resolution in the portions of *cyt aa₃* that are distant from the interface with ACIII suggests that there may be several conformations of the supercomplex that are all tethered by the loop in *cyt aa₃*. The loop could, therefore, serve as a hinge, enabling the membrane domains of ACIII and *cyt aa₃* to swing into contact transiently.

Using the location of TM3 to TM7 of subunit III within the supercomplex as a guide enables a model of the entire *cyt aa₃* to be placed within the density map for the supercomplex (Extended Data Fig. 9).

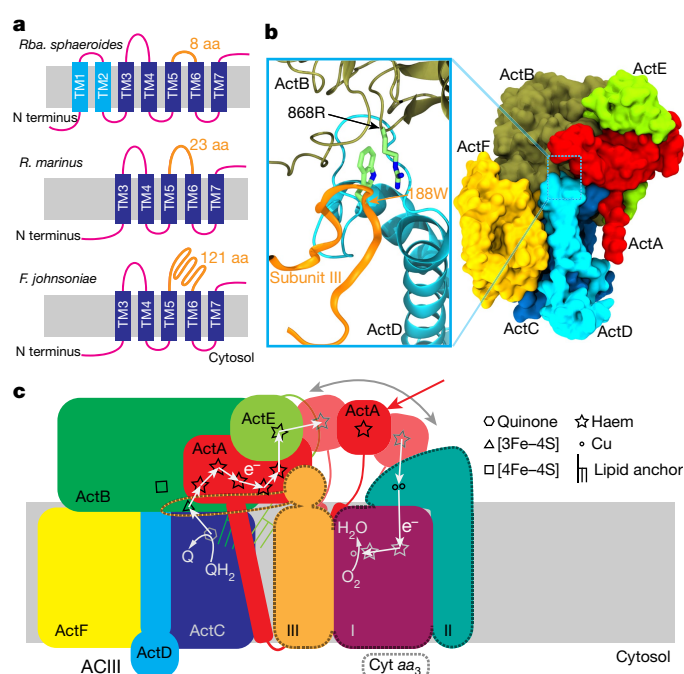


Fig. 4 | ACIII–*cyt aa₃* supercomplex in *F. johnsoniae*. **a**, Topology comparison of subunit III of *cyt aa₃* from different species. **b**, Tryptophan 188 from the subunit III loop (orange) of *cyt aa₃* interacts with arginine 868 from the subunit B of ACIII. The surface representation of ACIII on the right shows the binding pocket for the subunit III loop. **c**, Working model for electron transfer in the ACIII–*cyt aa₃* supercomplex.

In the resulting model, there is a considerable distance (56 Å) between the haem *c* in ActE of ACIII and Cu_A within subunit II of *cyt aa₃*. Electron transfer within the supercomplex does not require the addition of exogenous *cyt c*, which is also the case for the *cyt bcc*–*cyt aa₃* supercomplex from *C. glutamicum*²³. It is possible, although it seems unlikely, that there is a subset of conformations in which ActE comes close enough to *cyt aa₃* for direct electron transfer. It is noteworthy that the monohaem domain of ActA has substantial sequence homology (around 30% identity) with the haem *c* domain that is present at the C terminus of subunit II of *cyt caa₃* from *T. thermophilus*. This observation suggests that the ActA monohaem domain, which we postulate to be highly mobile in the structure (see above), may be able to interact with subunit II of *cyt aa₃* and shuttle electrons from the ActE monohaem domain to subunit II of *cyt aa₃*. As such, electron transfer within the supercomplex may require the monohaem domain of ACIII to swing back and forth between ACIII and *cyt aa₃* to shuttle electrons (Fig. 4c). Additional experimental work will be required to test this model and, indeed, to determine the physiological advantage of forming the supercomplex.

We would like to note that, contemporaneously with our studies, Sousa et al.²⁴ determined the structure of the homologous ACIII from *Rhodothermus marinus* by cryo-EM at 3.9 Å resolution. Aside from species-specific variations, the reported structures of the ACIIIs of *F. johnsoniae* and *R. marinus* are compatible. The observation made here that, like canonical complexes III and IV, ACIII and an adapted complex IV can also form a supercomplex hints at the importance of supercomplexes in oxidative phosphorylation. We demonstrate that high-resolution cryo-EM with SMA nanodiscs, which preserves native protein–protein and protein–lipid interactions, is ideally suited to have an important role in future studies.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at <https://doi.org/10.1038/s41586-018-0061-y>.

Received: 26 November 2017; Accepted: 20 March 2018;
Published online 25 April 2018.

- Pereira, M. M., Carita, J. N. & Teixeira, M. Membrane-bound electron transfer chain of the thermohalophilic bacterium *Rhodothermus marinus*: a novel multihemic cytochrome *bc*, a new complex III. *Biochemistry* **38**, 1268–1275 (1999).
- Yanyushin, M. F., del Rosario, M. C., Brune, D. C. & Blankenship, R. E. New class of bacterial membrane oxidoreductases. *Biochemistry* **44**, 10037–10045 (2005).
- Refojo, P. N., Sousa, F. L., Teixeira, M. & Pereira, M. M. The alternative complex III: a different architecture using known building modules. *Biochim. Biophys. Acta* **1797**, 1869–1876 (2010).
- Pereira, M. M., Refojo, P. N., Hreggvidsson, G. O., Hjorleifsdottir, S. & Teixeira, M. The alternative complex III from *Rhodothermus marinus* - a prototype of a new family of quinol:electron acceptor oxidoreductases. *FEBS Lett.* **581**, 4831–4835 (2007).
- Gao, X., Xin, Y., Bell, P. D., Wen, J. & Blankenship, R. E. Structural analysis of alternative complex III in the photosynthetic electron transfer chain of *Chloroflexus aurantiacus*. *Biochemistry* **49**, 6670–6679 (2010).
- Refojo, P. N., Ribeiro, M. A., Calisto, F., Teixeira, M. & Pereira, M. M. Structural composition of alternative complex III: variations on the same theme. *Biochim. Biophys. Acta* **1827**, 1378–1382 (2013).
- Dörr, J. M. et al. Crystal structure of the cytochrome *bc*₁ complex from bovine heart mitochondria. *Science* **277**, 60–66 (1997).
- Rothery, R. A., Workun, G. J. & Weiner, J. H. The prokaryotic complex iron–sulfur molybdoenzyme family. *Biochim. Biophys. Acta* **1778**, 1897–1929 (2008).
- Dietrich, W. & Klimmek, O. The function of methyl-menaquinone-6 and polysulfide reductase membrane anchor (PsrC) in polysulfide respiration of *Wolinella succinogenes*. *Eur. J. Biochem.* **269**, 1086–1095 (2002).
- Hedderich, R. et al. Anaerobic respiration with elemental sulfur and with disulfides. *FEMS Microbiol. Rev.* **22**, 353–381 (1998).
- Xia, D. et al. Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K⁺ channel: the power of native nanodiscs. *Proc. Natl Acad. Sci. USA* **111**, 18607–18612 (2014).
- Postis, V. et al. The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim. Biophys. Acta* **1848**, 496–501 (2015).
- Dörr, J. M. et al. The styrene–maleic acid copolymer: a versatile tool in membrane research. *Eur. Biophys. J.* **45**, 3–21 (2016).
- Parmar, M. et al. Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim. Biophys. Acta* **1860**, 378–383 (2018).
- Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into nanodiscs. *FEBS Lett.* **584**, 1721–1727 (2010).
- Jormakka, M. et al. Molecular mechanism of energy conservation in polysulfide respiration. *Nat. Struct. Mol. Biol.* **15**, 730–737 (2008).
- Majumder, E. L., King, J. D. & Blankenship, R. E. Alternative complex III from phototrophic bacteria and its electron acceptor auracyanin. *Biochim. Biophys. Acta* **1827**, 1383–1391 (2013).
- Refojo, P. N., Teixeira, M. & Pereira, M. M. The alternative complex III: properties and possible mechanisms for electron transfer and energy conservation. *Biochim. Biophys. Acta* **1817**, 1852–1859 (2012).
- Crofts, A. R. The cytochrome *bc*₁ complex: function in the context of structure. *Annu. Rev. Physiol.* **66**, 689–733 (2004).
- Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16**, 375–388 (2015).
- Kovacs-Simon, A., Titball, R. W. & Michell, S. L. Lipoproteins of bacterial pathogens. *Infect. Immun.* **79**, 548–561 (2011).
- Kao, W. C. et al. The obligate respiratory supercomplex from *Actinobacteria*. *Biochim. Biophys. Acta* **1857**, 1705–1714 (2016).
- Graf, S. et al. Rapid electron transfer within the III–IV supercomplex in *Corynebacterium glutamicum*. *Sci. Rep.* **6**, 34098 (2016).
- Sousa, J. S. et al. Structural basis for energy transduction by respiratory alternative complex III. *Nat. Commun.* (in the press).

Acknowledgements This work was supported by funds from the National Institutes of Health (R01-HL016101 to R.B.G.; P41-GM104601, U54-GM087519 and R01-GM123455 to E.T.) and Canadian Institutes of Health Research (MOP-81294 to J.L.R.); J.L.R. was supported by the Canada Research Chairs program. Some of this work was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy, supported by grants from the Simons Foundation (349247), and the National Institute of General Medical Sciences (P41-GM103310, S10-OD019994). Molecular dynamics simulations were performed at Blue Waters (ACI-1713784 to E.T.) and XSEDE (TG-MCA06N060 to E.T.). Blue Waters is supported by the National Science Foundation (OCI-0725070 and ACI-1238993) and the State of Illinois. XSEDE is supported by the National Science Foundation (ACI-1548562). We thank M. McBride for providing us with the *Flavobacterium johnsoniae* UW101 strain, B. Carragher and C. Potter for facilitating access to the Titan Krios, Z. Zhang for collecting data, and M. Mazhab-Jafari for advice on building atomic models.

Reviewer information *Nature* thanks Y. Cheng, M. Wikström and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions P.V. expressed, purified and characterized the protein supercomplex, S.B. prepared cryo-EM specimens and calculated cryo-EM maps, C.S. built the de novo atomic structures, Y.W. conducted the molecular dynamics simulations, S.H. performed the electrochemical characterization, J.H. performed the metal analysis, E.T. supervised molecular dynamics simulations, J.L.R. supervised the molecular structure determination, and R.B.G. conceived the project, supervised biochemical experiments, and coordinated the project. All authors participated in the manuscript preparation.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-018-0061-y>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-018-0061-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to E.T., J.L.R. or R.B.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Bacterial strain and growth conditions. *Flavobacterium johnsoniae* ATCC 17061 strain UW101 was used in this study. The strain was a gift from M. McBride at the University of Wisconsin, Milwaukee. The cells were grown in casitone-yeast extract medium at 30 °C under high aerobic conditions (500 ml cultures in 2 l flasks)²⁵.

Membrane preparation and protein purification. Cells grown overnight were collected by centrifugation (14,000g for 10 min). The cell pellet from 12 l of culture (~2.5 g l⁻¹) was resuspended in ~200 ml of 20 mM Tris-HCl buffer, pH 8 (buffer A) with 5 mM MgSO₄, DNase I (Sigma) and a protease inhibitor cocktail (Sigma). This suspension was passed three times through a Microfluidizer at a pressure of 80,000 psi to disrupt the cells. The cell extract was centrifuged at 14,000g for 10 min to remove unbroken cells. Membranes were obtained after centrifugation at 185,500g for 4 h. Under the above growth conditions, the membranes contained ACIII, cyt *aa*₃ and cyt *bd*. The membrane pellet was solubilized by using either a traditional detergent or the SMA copolymer.

Purification using Triton X-100 and DDM. The membrane pellet was resuspended in buffer A (~50 mg ml⁻¹) along with 300 mM NaCl, and solubilized by the addition of Triton X-100 (Fisher Scientific) to a final concentration of 4%. The solution was incubated at 4 °C for 2 h with mild agitation. The suspension was cleared by centrifugation at 185,500g for 1 h, after which the detergent was diluted fourfold by adding three volumes of buffer A to the supernatant. The diluted supernatant was then added to a chromatography column containing 10 ml of Ni-NTA resin (Qiagen) pre-equilibrated with 20 mM Tris-HCl pH 8 containing 0.05% Triton X-100 and 0.15 M NaCl (buffer B). The resin was washed with about ten column volumes of buffer B to remove any unbound sample. Detergent exchange to *n*-dodecyl-β-D-maltoside (DDM; Anatrace) was carried out by washing with buffer B containing 0.05% DDM instead of Triton X-100 (buffer C). The column was further washed with five column volumes of buffer C containing 10 mM imidazole to remove the loosely bound proteins from the resin. The proteins that were well bound to the resin were eluted using 100 mM imidazole in buffer C. The eluent was concentrated to around 3 ml using Amicon Ultra-15 filters (Millipore) with a 100-kDa cutoff. The excess imidazole was removed by dialysis against buffer C. The yield of protein obtained was about 0.3 mg l⁻¹ of ACIII and 0.16 mg l⁻¹ of cyt *aa*₃ from 12 l of culture. When indicated, the proteins were further purified by gel filtration chromatography using a Superdex 200 10/300 GL column (GE Healthcare Life Sciences). The purified proteins were stored at -80 °C after adding glycerol to a final concentration of 10%.

Purification using SMA copolymer. The SMA copolymer SMA 3000HNA (styrene maleic acid copolymer, ~3:1 molar ratio of styrene:maelic acid) was a gift from T. Bricker (Louisiana State University) who used SMA copolymer made by Cray Valley USA (now Total Petrochemicals & Refining USA) successfully for the studies of photosystem from spinach thylakoids²⁶. Additional SMA 3000HNA was provided by Total Petrochemicals & Refining USA as an aqueous solution of 25.6% (w/v) SMA. We also used a similar product, Xiran SL25010 S25, provided by Polyscope Polymers B. V., with similar results. These polymer preparations are provided as aqueous solutions of the sodium salt, and the polymer solutions were simply diluted to the final desired percentage to use directly for the solubilization of membranes. The purification protocol with the SMA copolymer was similar to that described with detergents with the following differences. After the membrane pellet was resuspended, the SMA solution was added dropwise to a final concentration of 1% with continuous stirring. After incubation for 1 h at room temperature, the solution was centrifuged at 185,500g for 1 h to remove unsolubilized particles. The supernatant was loaded directly to the Ni-NTA column equilibrated with 20 mM Tris-HCl pH 8, 0.15 M NaCl. The remaining steps of the purification were as described above. After solubilization of the membrane suspension with 1% SMA 3000HNA, no additional SMA or detergents were added and were not needed to maintain the solubilized proteins in solution. The yield of protein after the use of the SMA copolymer for solubilization was about 0.5 mg l⁻¹ for ACIII and about 0.15 mg l⁻¹ for cyt *aa*₃ from 12 l of culture.

Analytical methods. The total protein concentration was determined using the BCA kit (Thermo Scientific, Pierce Protein Research Products). The UV-visible absorption spectra of the oxidized and reduced proteins were recorded on an Agilent Technologies spectrophotometer (model 8453). The pyridine haemochrome assay²⁷ was used to determine the concentration of haems present in the protein samples. The total haem *c* concentration was divided by seven to calculate the ACIII concentration and the total haem *a* concentration was divided by two to calculate the cyt *aa*₃ concentration. The purified protein was analysed by SDS-PAGE using 4–20% precast gels (Nusep Tech). Haem staining was carried out using 3,3',5,5'-tetramethyl benzidine (TMBZ)²⁸. The supercomplex was visualized by blue native PAGE (BN-PAGE) using a 4–16% gel (Novex, Life Technologies) with Bis-Tris buffer. The entire gel was stained with Coomassie blue, and then fixed with 30% methanol and 10% glacial acetic acid. The gel was destained with 8% glacial acetic acid to visualize the bands. Peptide mass spectrometry and analyses were carried out by P. Yau at

the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign.

Oxygen consumption assay. Oxygen consumption was measured using a Clark electrode (Strathkelvin) in a 1 ml chamber at 25 °C as previously described²⁹. The reaction mix consisted of 100 μM ubiquinone-1 (Q₁; Sigma-Aldrich) and 5 mM dithiothreitol in air-saturated 0.1 M potassium phosphate buffer, pH 7.5 with 150 mM NaCl. The reaction was started by adding the purified protein into the chamber. The initial concentration of oxygen was calculated to be 237 μM.

Quinol:cytochrome *c* oxidoreductase activity. The quinol:cytochrome *c* oxidoreductase activity of the ACIII was measured spectrophotometrically as described previously³⁰. The reaction was carried out in a 2 ml anaerobic cuvette, at 25 °C in 50 mM potassium phosphate buffer, pH 7.5 in the presence of 50 μM horse heart cytochrome *c* (Sigma-Aldrich) and 200 μM KCN. Ubiquinol-1 (Q₁H₂) or reduced vitamin K₂ (Sigma-Aldrich) were used as quinol substrates and, in each case, the quinone was reduced using sodium borohydride according to a previously described method³¹. The reaction was started by the addition of 100 μM of reduced quinone.

EPR spectroscopy. The purified ACIII-cyt *aa*₃ supercomplex was extensively dialysed against 20 mM Tris-HCl buffer, pH 8, with 150 mM NaCl and 1 mM EDTA to eliminate adventitious transition metal ions. The sample was concentrated in an Amicon filter to 150 μl with a final ACIII concentration of around 60 μM. The air-oxidized sample was directly transferred to an X-band EPR tube and subsequently frozen in liquid nitrogen. The sample was oxidized completely by the addition of 2 mM potassium ferricyanide. Glycerol (5%) was present in all EPR samples. Continuous wave EPR measurements were carried out on an X-band Varian EPR-E122 spectrometer at the Electron Paramagnetic Resonance facility at the University of Illinois at Urbana-Champaign. Cryogenic conditions below 77 K were achieved with a Lakeshore 331 temperature controller using a regulated flow of helium gas.

Metal analysis. Metal analysis was carried out using inductively coupled plasma mass spectrometry (ICP-MS) as previously described^{32,33}.

Optical redox titration. Full spectrum UV-visible redox titrations were performed to determine the midpoint potentials (E_m^o) of the redox-active cytochromes in the DDM-solubilized ACIII-cyt *aa*₃ supercomplex^{34,35}. The purified supercomplex was suspended in 4 ml of 50 mM potassium phosphate buffer pH 7.0 to a concentration of 3 μM with 25 μM each of the following redox mediators: benzyl viologen ($E_{m,7} = -350$ mV), anthraquinone-2-sulfonate ($E_{m,7} = -225$ mV), 2-hydroxy-1,4-naphthoquinone ($E_{m,7} = -220$ mV), 9,10-anthraquinone-2,6-disulfonate ($E_{m,7} = -185$ mV), duroquinone ($E_{m,7} = 5$ mV), *N*-ethylphenazonium ethosulfate ($E_{m,7} = 65$ mV), *N*-methylphenazonium methosulfate ($E_{m,7} = 85$ mV), diaminodurene ($E_{m,7} = 275$ mV), 2,6-dimethyl benzoquinone ($E_{m,7} = 180$ mV), 1,2-naphthoquinone ($E_{m,7} = 143$ mV), 1,4-naphthoquinone ($E_{m,7} = 36$ mV) and potassium ferricyanide ($E_{m,7} = 435$ mV)³⁶. Titrations were performed with an anaerobic stirred cuvette and the solution potential was adjusted by injecting aliquots of 10 mM sodium dithionite or potassium ferricyanide as reductant and oxidant, respectively. Spectra were taken at approximately 10–20 mV increments over the titration range indicated. Spectroscopic changes of the α-bands of the haems upon reduction or oxidation were monitored at the peak maxima to determine the midpoint potentials of each class of haem centre. The datasets were analysed using Origin (Origin Laboratory Corporation) to determine spectral components and fit titration curves using the Nernst equation³⁶.

Electron microscopy sample preparation. Holey carbon film-coated electron microscopy grids were nanofabricated with regular arrays of 500- to 800-nm holes³⁷ and coated with an additional layer of gold. Cryo-EM specimens were prepared with a FEI Vitrobot grid preparation robot at 4 °C and 100% humidity by applying 3 μl of sample (3 mg ml⁻¹) to glow-discharged grids, allowing the grids to equilibrate for 1 s, and blotting for 12 s before freezing in a liquid ethane:propane mixture (1:1 v/v)³⁸. Grids were subsequently stored in liquid nitrogen before shipping to the New York Structural Biology Center for imaging with a FEI Titan Krios electron microscope equipped with a Gatan K2 Summit camera and automated with Leginon³⁹.

Electron microscopy data acquisition. Movies were acquired in electron counting mode with a pixel size of 1.1 Å, an exposure rate of 7.4 electrons per pixel per second, and a total exposure time of 10 s divided in 40 frames (418 movies) or 50 frames (1,599 movies). Frame alignment and exposure weighting were performed with Motioncor2⁴⁰. After screening averages from the aligned movies, 475 movies were discarded because of excessive movement, low defocus, high defocus, or overfocus. Contrast transfer function parameters were estimated from the exposure-weighted averages of movie frames with CTFFIND4⁴¹.

Image processing. Particle images (3,044) were manually selected and subjected to 2D classification with Relion 1.4⁴². The resulting 2D classes were used as templates for automatic selection of 899,405 particle images⁴³. The number of particle images was reduced to 693,416 by further 2D classification. Subsequent image processing was carried out in cryoSPARC⁴⁴. An initial map of ACIII-cyt *aa*₃ was obtained

by ab initio 3D classification, refined to 4.1 Å resolution, and used as a reference for the multi-refine procedure in cryoSPARC producing initial maps of the ACIII and the ACIII-cyt *aa*₃. Particle images (164,239) were used to refine the ACIII-cyt *aa*₃ map to 3.4 Å resolution, but this map showed the cyt *aa*₃ portion of the complex with lower density than the ACIII part. Maps with uniform density for ACIII-cyt *aa*₃ and ACIII, both at 3.6 Å resolution, were calculated from 81,530 and 51,547 particle images, respectively.

Model building. The 3.4 Å resolution density map was used for the de novo model building of ACIII. The density map was first segmented with UCSF Chimera⁴⁵ to facilitate the identification of subunits. The connectivity of each segmented map was further examined and the result was compared with topology predictions from topocons⁴⁶ and secondary structure prediction from Jpred⁴⁷ to validate the subunit assignment and identify the directionality of peptide chain. With this information, model building was carried out manually in Coot⁴⁸. Individual chains were first traced in C α baton mode. Readily interpretable features from the density map, including regions rich in bulky residues, triacylated cysteines, and axial ligands of haem c, were used to register the structure to the sequence. Stretches of ~20 amino acids were built progressively around these registration points and assembled as a single chain in Coot. All six subunits of ACIII were combined and refined with phenix.real_space_refine⁴⁹. For cofactors, the starting models were taken from the CCP4 ligand library directly. Cofactors were docked to the density map with Coot and merged with the apo protein structure. The complete structure was then refined with phenix.real_space_refine with geometric constraints for the protein-cofactor coordination. The final model was further examined in Coot to remove amino acid side chains with ambiguous orientations and further validated with MolProbity⁵⁰ and EMRinger⁵¹. All identified lipids with two acyl tails were modelled as phosphatidylethanolamine with palmitoyl tail. The conformation of phosphatidylethanolamine was refined with interactive molecular dynamics flexible fitting (iMDFF) in the presence of the protein structure using VMD⁵². Lipid tails were then truncated according to the density map.

The 3.6 Å resolution ACIII-cyt *aa*₃ density map was used for the model building of cyt *aa*₃. Part of the subunit III loop region was manually built in Coot. Homology models for individual subunits were generated with the RaptorX server⁵³ and docked into the density map with UCSF Chimera. The model for the ACIII-cyt *aa*₃ supercomplex was assembled by fitting the ACIII structure to the ACIII-cyt *aa*₃ map and placing the cyt *aa*₃ structure from *Rba. sphaeroides* (PDB 1M56) into the map based on the position of cyt *aa*₃ subunit III.

Bioinformatic analysis. Homologous protein sequences were retrieved using the NCBI blastp server⁵⁴. The blastp results were analysed in python 2.7 with pandas and biopython modules. Sequence hits were filtered on the basis of coverage and sequence identity. Representative sequences were selected on the basis of sequence identity to maintain the variations in sequence and aligned using the Clustal Omega server⁵⁵. Figures for sequence alignment were prepared using the ESPript 3.0 server⁵⁶.

Simulation system preparation. The initial ACIII structure for the MD simulation was obtained from the refined structure determined by cryo-EM. Eleven 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipids resolved by cryo-EM were added to the ACIII system, which was subsequently embedded in a POPE membrane bilayer, solvated with the TIP3P water model⁵⁷, and ionized with 150 mM NaCl.

ReMDFF simulation. Resolution-exchange molecular dynamics flexible fitting (ReMDFF)⁵⁸ was used for structure refinement, with the CHARMM36m force field for proteins⁵⁹ and the CHARMM36 force field for lipids⁶⁰. Force-field parameters for haems and iron-sulfur clusters came from previous studies^{61–63}. The fitting was performed in vacuum in the presence of a grid potential derived from the experimental density map (coupling factor 0.3). Secondary structure restraints, *cis*-peptide bond restraints, and chirality restraints were applied to the protein. Haems and iron-sulfur clusters were harmonically restrained ($k = 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). A Langevin thermostat⁶⁴ was used for maintaining the average temperature at 80 K. The MD integration time step was 1 fs. A cut-off radius for nonbonded interactions was set to 10 Å with a switching function taking effect at 9 Å. A total of six replicas were used together with six grid potentials of decreasing resolution. Each was first energy-minimized for 2,000 steps and then equilibrated for 1 ps. Finally, 2,000 replica exchanges were attempted with 1 ps between attempts.

Molecular dynamics simulation. The ACIII systems were simulated with NAMD 2.12 using the same force-field parameters as in ReMDFF. The system was energy-minimized for 3,000 steps using the conjugated gradient algorithm⁶⁵ with linear searching⁶⁶, and equilibrated for 0.5 ns to relax lipid tail group atoms while keeping the lipid phosphorus atoms and protein (including haems and iron-sulfur clusters) heavy atoms harmonically restrained ($k = 1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). This procedure was followed by a 10-ns simulation to allow lipids to relax around the proteins while keeping the protein backbone and heavy atoms from iron-sulfur clusters and haems harmonically restrained ($k = 1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). Restraints were gradually

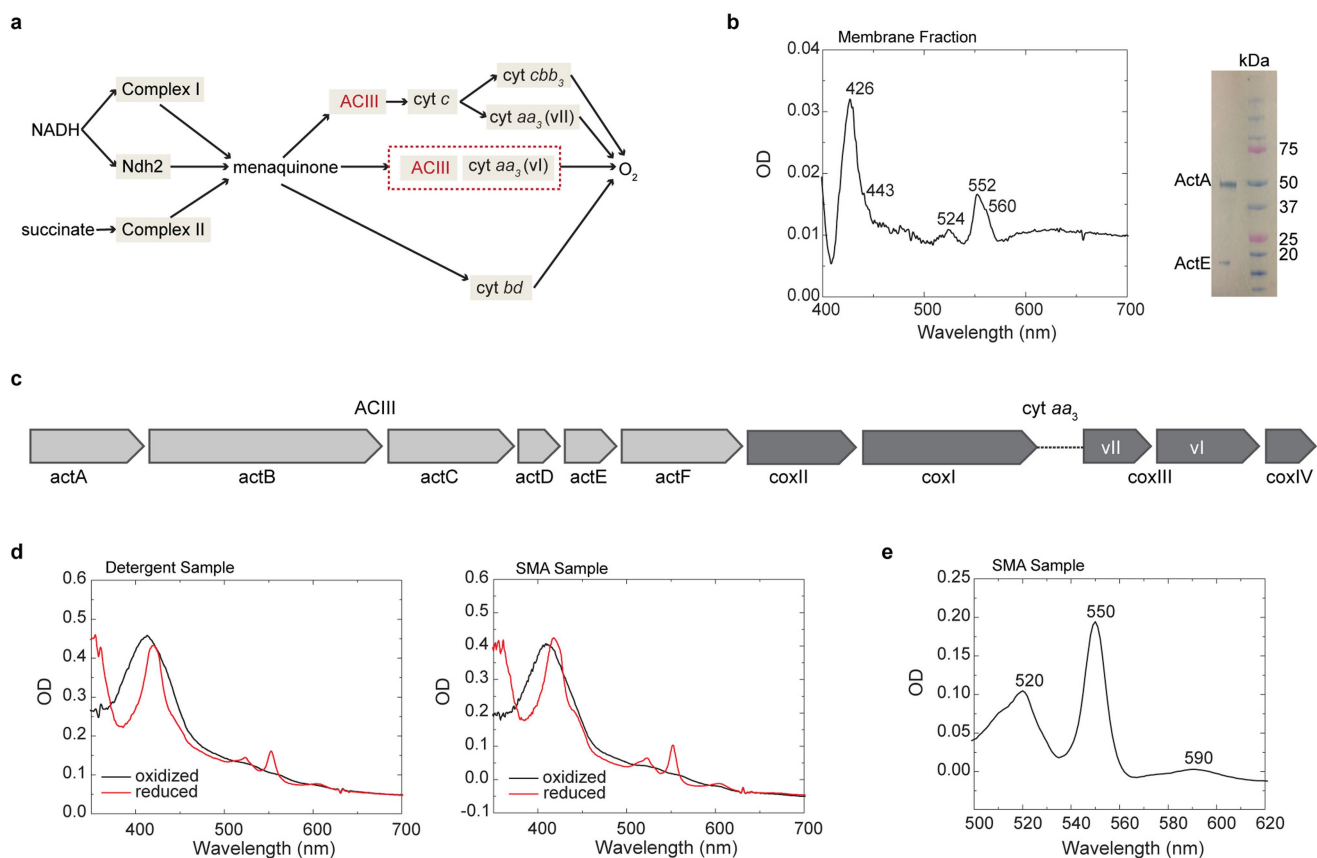
released over the next 5 ns and the simulation continued without any biasing potential for a total of 250 ns. The angles in the iron-sulfur clusters were harmonically restrained to their initial values ($k = 300 \text{ kcal mol}^{-1} \text{ deg}^{-1}$) throughout the simulation.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. All relevant data are included in the manuscript or Supplementary Information and/or are available from the corresponding authors upon reasonable request. Three cryo-EM maps mentioned in this work have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-7286 (combined), EMD-7447 (ACIII-cyt *aa*₃), EMD-7448 (ACIII). The coordinates of the atomic model of the alternative complex III built from EMD-7286 have been deposited in the Protein Data Bank (PDB) under accession code 6BTM.

- McBride, M. J. & Baker, S. A. Development of techniques to genetically manipulate members of the genera *Cytophaga*, *Flavobacterium*, *Flexibacter*, and *Sporocytophaga*. *Appl. Environ. Microbiol.* **62**, 3017–3022 (1996).
- Bell, A. J., Frankel, L. K. & Bricker, T. M. High yield non-detergent isolation of photosystem I-light-harvesting chlorophyll II membranes from spinach thylakoids: implications for the organization of the PS I antennae in higher plants. *J. Biol. Chem.* **290**, 18429–18437 (2015).
- Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).
- Thomas, P. E., Ryan, D. & Levin, W. An improved staining procedure for the detection of the peroxidase activity of cytochrome P-450 on sodium dodecyl sulfate polyacrylamide gels. *Anal. Biochem.* **75**, 168–176 (1976).
- Minghetti, K. C. et al. Modified, large-scale purification of the cytochrome o complex (*bo*-type oxidase) of *Escherichia coli* yields a two heme/one copper terminal oxidase with high specific activity. *Biochemistry* **31**, 6917–6924 (1992).
- Gao, X., Xin, Y. & Blankenship, R. E. Enzymatic activity of the alternative complex III as a menaquinol:auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. *FEBS Lett.* **583**, 3275–3279 (2009).
- Ragan, C. I., Wilson, M. T., Darley-Usmar, V. M. & Lowe, P. N. *Mitochondria: A Practical Approach* (IRL Press, Oxford, 1987).
- Carrell, C. J. et al. Generation of novel copper sites by mutation of the axial ligand of amicyanin. Atomic resolution structures and spectroscopic properties. *Biochemistry* **46**, 1900–1912 (2007).
- Ouyang, H. et al. Functional importance of a pair of conserved glutamic acid residues and of Ca²⁺ binding in the *cbb*₃-type oxygen reductases from *Rhodobacter sphaeroides* and *Vibrio cholerae*. *Biochemistry* **51**, 7290–7296 (2012).
- Bowyer, J. R., Tierney, G. V. & Crofts, A. R. Cytochrome *c*₂ – reaction centre coupling in chromatophores of *Rhodospseudomonas sphaeroides* and *Rhodospseudomonas capsulata*. *FEBS Lett.* **101**, 207–212 (1979).
- Dutton, P. L. Redox potentiometry: determination of midpoint potentials of oxidation-reduction components of biological electron-transfer systems. *Methods Enzymol.* **54**, 411–435 (1978).
- Fultz, M. & Durs, R. Mediator compounds for the electrochemical study of biological redox systems: a compilation. *Anal. Chim. Acta* **140**, 1–18 (1982).
- Marr, C. R., Benlekber, S. & Rubinstein, J. L. Fabrication of carbon films with ~500 nm holes for cryo-EM with a direct detector device. *J. Struct. Biol.* **185**, 42–47 (2014).
- Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
- Suloway, C. et al. Automated molecular microscopy: the new Legion system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
- Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
- Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407 (2015).
- Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Barad, B. A. et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

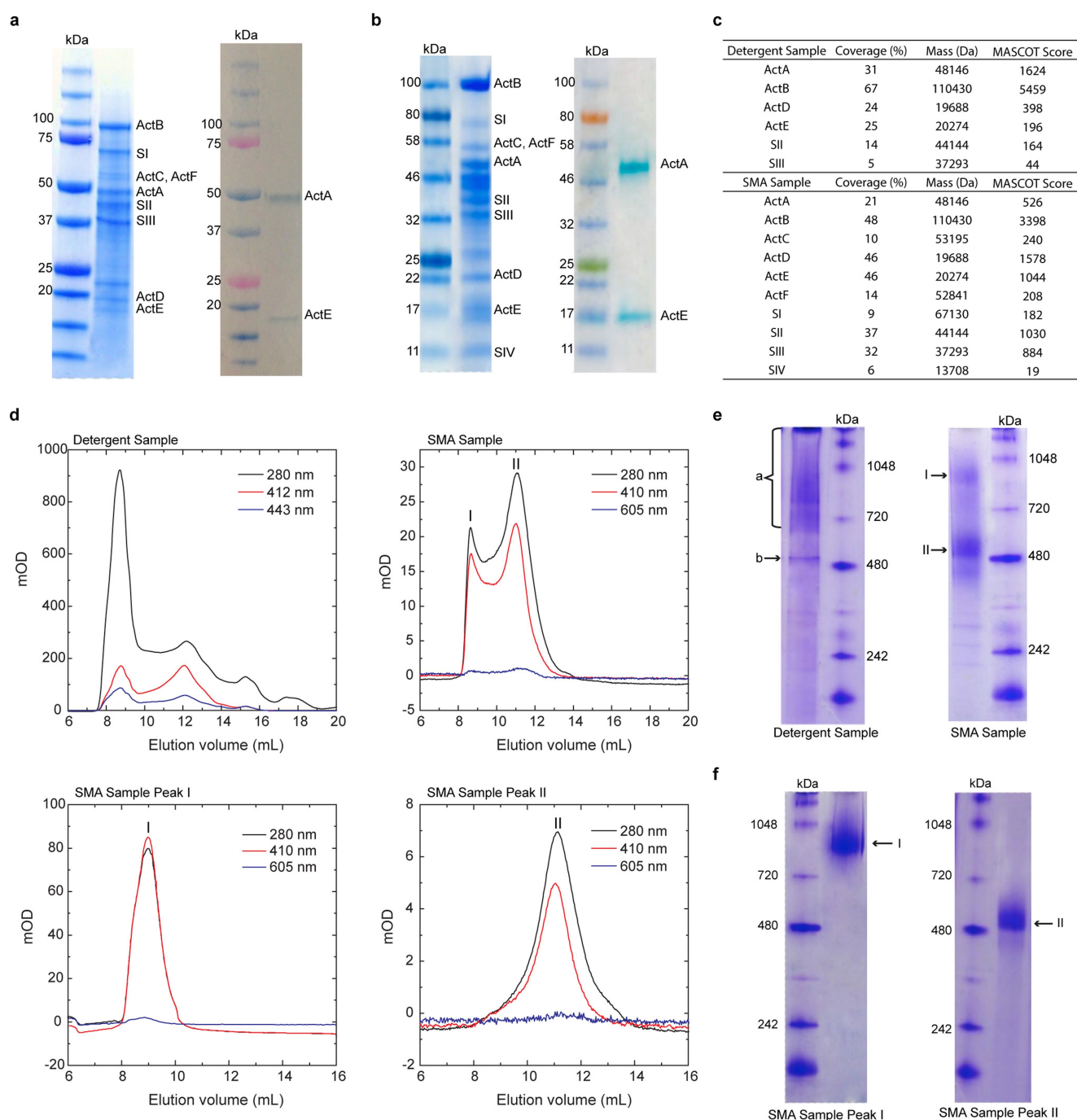
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
53. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
54. Madden, T. L., Tatusov, R. L. & Zhang, J. Applications of network BLAST server. *Methods Enzymol.* **266**, 131–141 (1996).
55. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
56. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
57. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
58. Singharoy, A. et al. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *eLife* **5**, 61–67 (2016).
59. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
60. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
61. Autenrieth, F., Tajkhorshid, E., Baudry, J. & Luthey-Schulten, Z. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.* **25**, 1613–1622 (2004).
62. Autenrieth, F., Tajkhorshid, E., Klaus Schulten, A. & Luthey-Schulten, Z. Role of water in transient cytochrome c_2 docking. *J. Phys. Chem. B* **108**, 20376–20387 (2004).
63. Carvalho, A. T. P. & Swart, M. Electronic structure investigation and parametrization of biologically relevant iron–sulfur clusters. *J. Chem. Inf. Model.* **54**, 613–620 (2014). erratum 55, 1508–1508 (2015).
64. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford Univ. Press, New York, 1987).
65. Fletcher, R. & Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **7**, 149–154 (1964).
66. Sun, W. & Yuan, Y.-X. *Optimization Theory and Methods: Nonlinear Programming* (Springer, New York, 2006).



Extended Data Fig. 1 | Expression and spectroscopic characterization of the ACIII-cyt *aa₃* supercomplex.

a, A schematic of the respiratory chain of *F. johnsoniae*. **b**, UV-visible spectrum and SDS-PAGE of the membranes from *F. johnsoniae*. Left, the difference spectrum of the membranes of *F. johnsoniae*, obtained from the spectrum of the air-oxidized membranes and the spectrum after reduction with dithionite. The wavelengths associated with the haem peaks are 605 nm, 560 nm, 552 nm and a broad peak at 630 nm for haems *a*, *b*, *c* and *d*, respectively. Right, the SDS-PAGE with the membranes followed by staining the gel for haems shows bands corresponding to the cytochrome subunits ActA (48 kDa) and ActE (20 kDa) of ACIII but no bands corresponding to the cytochrome subunit (around 35 kDa) from the *cbb₃* oxidase. **c**, The gene arrangement for the ACIII and the cytochrome oxidase *aa₃* genes in the *F. johnsoniae* genome. The genes for the subunits I and II from cyt *aa₃* oxidase are found immediately downstream of those for the *act*

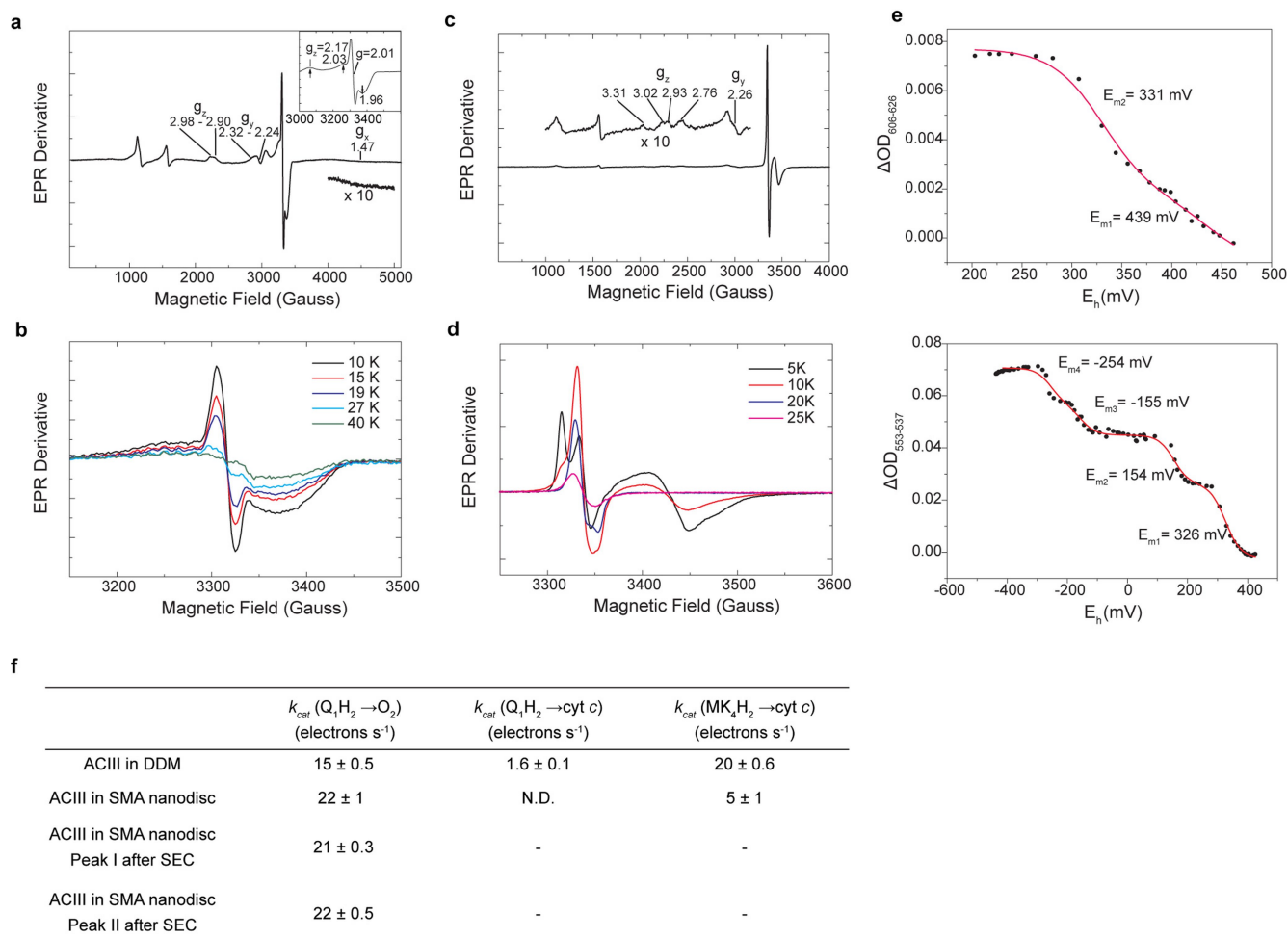
genes of the ACIII. Two different versions of subunit III are denoted as vI and vII. **d**, UV-visible spectra of the reduced and oxidized forms of the supercomplex in detergent and SMA nanodiscs. The dithionite reduced form of the samples is represented in red and shows the peaks for haem *c* at 524 nm and 552 nm and those for haem *a* at 443 nm and 605 nm. **e**, Pyridine haemochrome assay of the ACIII-cyt *aa₃* supercomplex in SMA nanodiscs. Plotted is the reduced-minus-oxidized difference spectrum of the pyridine haemochromes of the sample. Peaks at 520 nm and 550 nm are associated with haem *c* and the peak at 590 nm is associated with haem *a*. Quantification from the spectrum shows a ratio of 10.6:1 between haem *c* and haem *a*, which translates into a 3:2 ratio between ACIII and cyt *aa₃* assuming 7 haem *c* per ACIII and 2 haem *a* per cyt *aa₃*. Data in **b** are representative of two independent experiments with similar results, and data in **d** and **e** are representative of six independent experiments with similar results.



Extended Data Fig. 2 | Component and size analysis of ACIII-cyt aa_3 supercomplex. **a**, SDS-PAGE of the detergent-solubilized preparation followed by Coomassie staining (left) and haem staining (right).

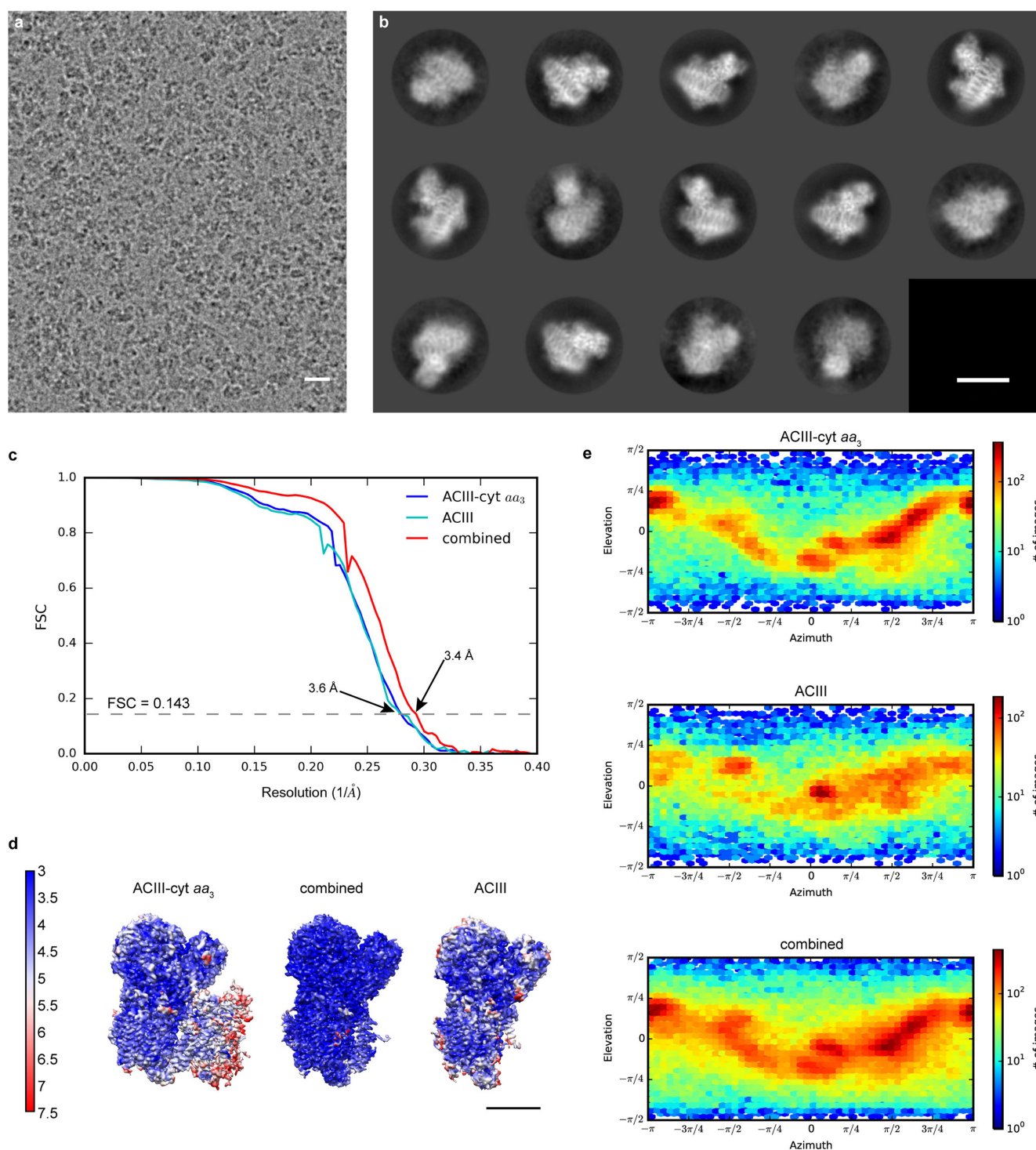
b, SDS-PAGE of the SMA nanodiscs preparation followed by Coomassie staining (left) and haem staining (right). **c**, Mass spectrometry results for the ACIII-cyt aa_3 supercomplex preparations. **d**, Size-exclusion chromatography with the ACIII-cyt aa_3 supercomplex from *F. johnsoniae*. Top left, the chromatogram of the detergent-solubilized sample, showing traces for protein at 280 nm, haem *c* at 412 nm and haem *a* at 443 nm respectively. Top right, the chromatogram of the sample isolated using the SMA copolymer, showing traces for protein at 280 nm, haem *c* at 410 nm and haem *a* at 605 nm. I and II are the two peaks corresponding to

two populations of the supercomplex. Bottom left, chromatogram of the fraction containing peak I. Bottom right, chromatogram of the fraction containing peak II. **e**, BN-PAGE of the ACIII-cyt aa_3 supercomplex. Left, the detergent-solubilized ACIII-cyt aa_3 supercomplex, showing a band at around 500 kDa, a smear of possible aggregates and possibly ACIII by itself. Right, the supercomplex in SMA nanodiscs, showing two different populations. **f**, BN-PAGE with the two different populations of ACIII-cyt aa_3 supercomplex in SMA nanodiscs purified from size-exclusion chromatography. The two chromatographic peaks correspond to the two bands observed in the BN-PAGE. Data in **a**, **b** are representative of six independent experiments and those in **d**–**f** are representative of three independent experiments with similar results.



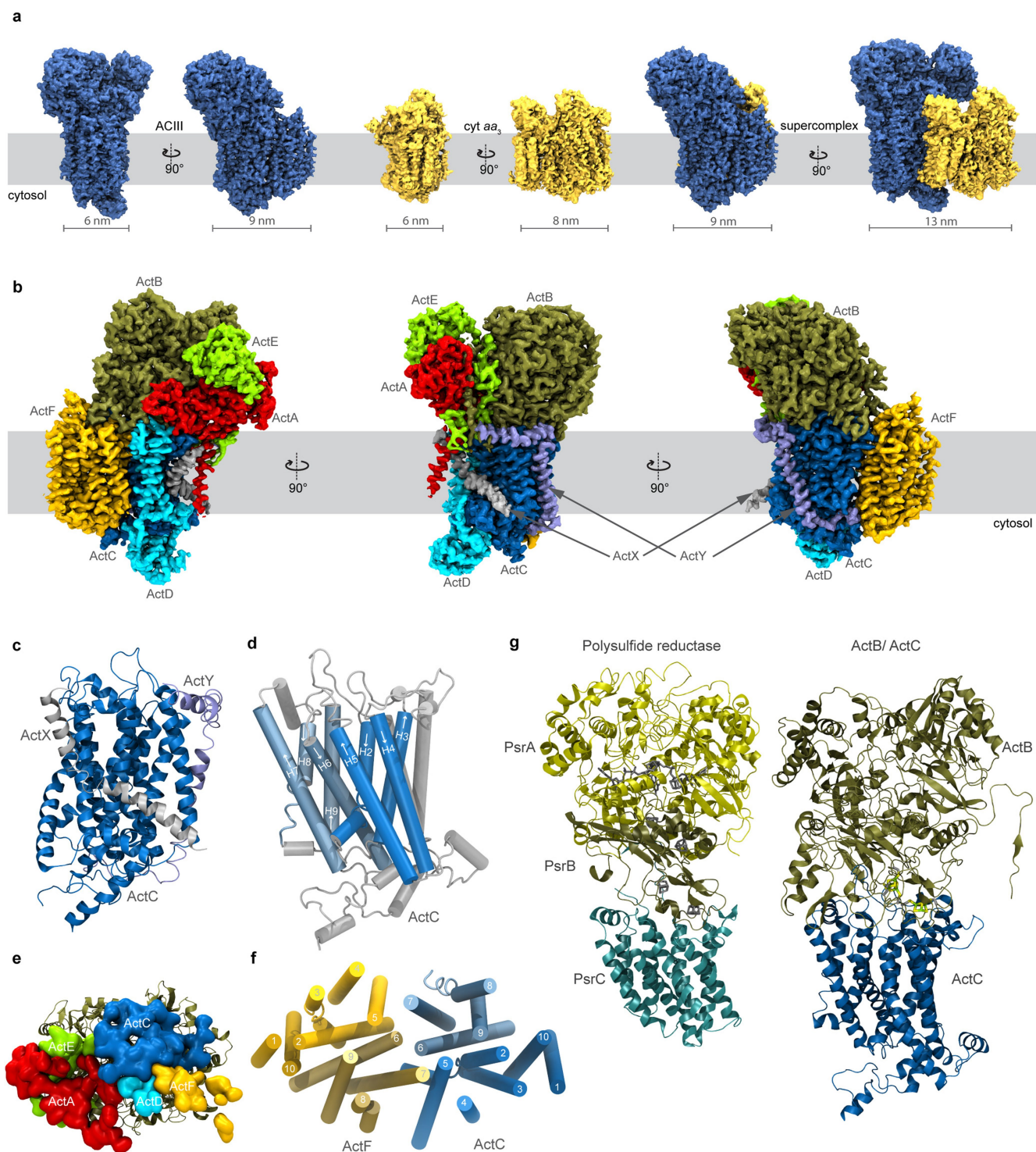
Extended Data Fig. 3 | Functional assays of the ACIII-cyt aa_3 supercomplex. **a**, The EPR spectrum of the air-oxidized sample showing peaks of the $[3\text{Fe}-4\text{S}]^+$ cluster from ACIII, the Cu_A from the cyt aa_3 oxidase and low-spin haems with overlapping g values. Insert is a zoomed view from 3,000 G to 3,500 G to better visualize the peaks from Cu_A (black arrows) and the $[3\text{Fe}-4\text{S}]^+$ cluster. The region between 4,000 G and 5,000 G is magnified ten times to show the broad g_x trough of low-spin haems. The measurement condition is 10 K, 9.267 GHz, 2 mW microwave power and 20 Gauss modulation. **b**, The EPR spectra of the ferricyanide-oxidized sample at various temperatures. The measurement condition is 9.257 GHz, 2 mW microwave power and 5 Gauss modulation. **c**, The EPR spectrum of the air-oxidized sample showing peaks of iron-sulfur clusters from ACIII

and low-spin haems. The measurement condition is 10 K, 9.427 GHz, 2 mW microwave power, 10 Gauss modulation. **d**, The EPR spectra of the air-oxidized sample at various temperatures. The measurement condition is 9.427 GHz, 2 mW microwave power, 5 Gauss modulation. **e**, Redox titration of the haems in the ACIII and the cyt aa_3 oxidase in supercomplex in DDM. The potentiometric titration of the c haems from the ACIII (top) and the a haems from the cyt aa_3 oxidase (bottom). The E_m values are indicated and the solid red line represents the Nernst fitting. **f**, Steady-state activity of the ACIII-cyt aa_3 preparations. The number of independent experiments is six for ACIII in DDM and SMA nanodiscs, and three for peak I and peak II. Data are means \pm s.d. Data in a–e are representative of three independent experiments with similar results.



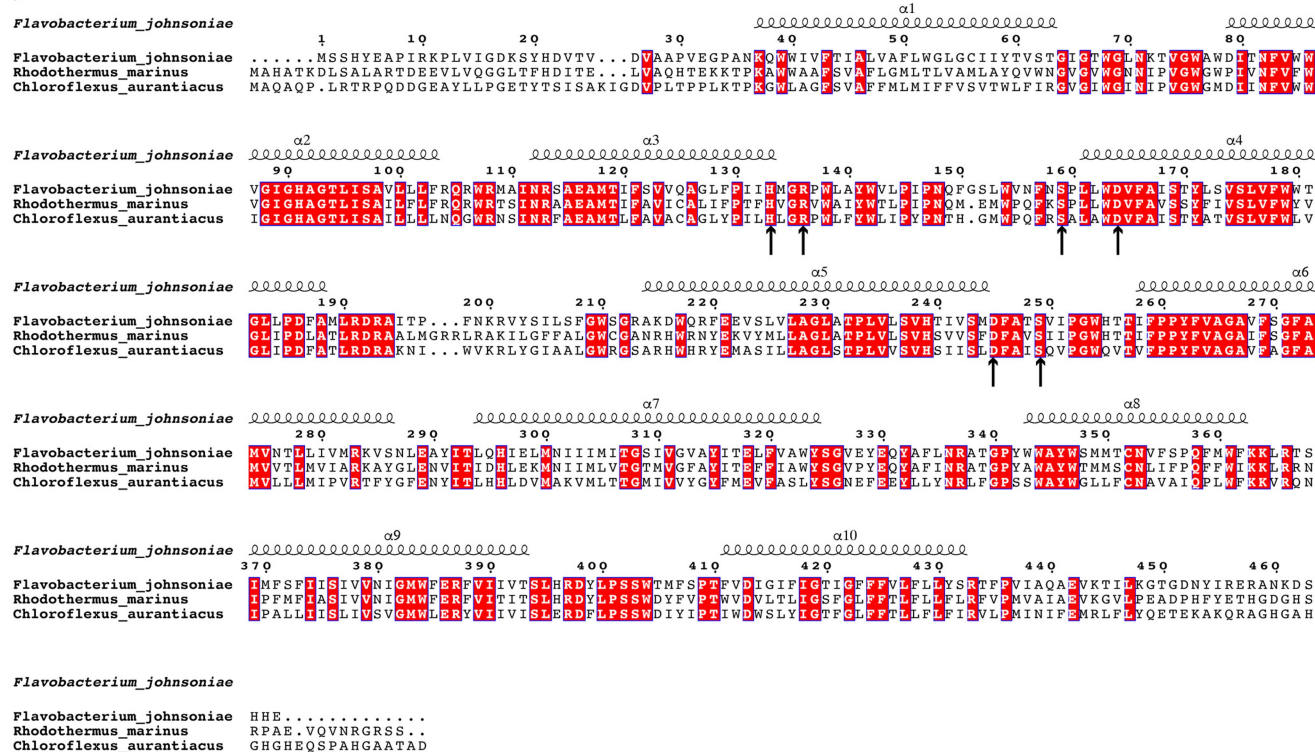
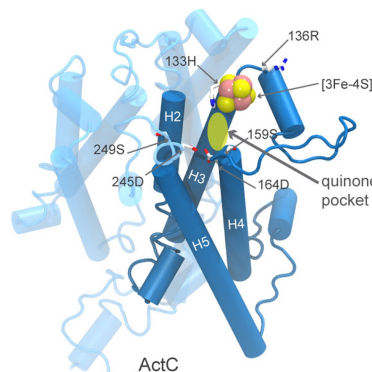
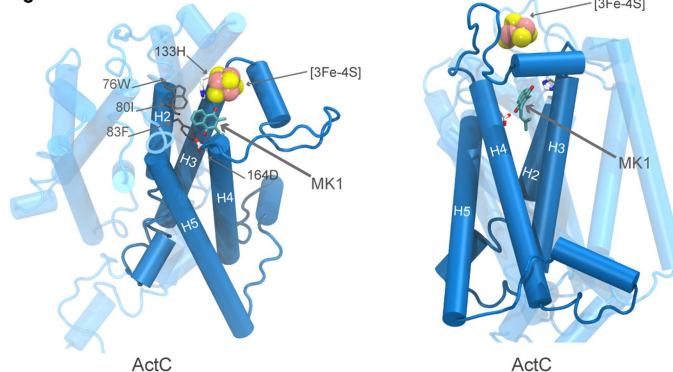
Extended Data Fig. 4 | Single-particle cryo-EM of the ACIII-cyt aa_3 supercomplex in SMA nanodiscs. **a**, Sum of an aligned movie of the ACIII-cyt aa_3 supercomplex in an SMA nanodisc. Scale bar, 20 nm. **b**, Two-dimensional class averages. Scale bar, 10 nm. **c**, Fourier shell coefficient curves between two independently refined half-maps for the

ACIII-cyt aa_3 map, ACIII map and combined map. **d**, Surface rendering maps coloured according to local resolution. Scale bar, 5 nm. **e**, Euler angle distributions of particles included in the calculation of the three final maps. Data collection and structure calculation were not repeated.



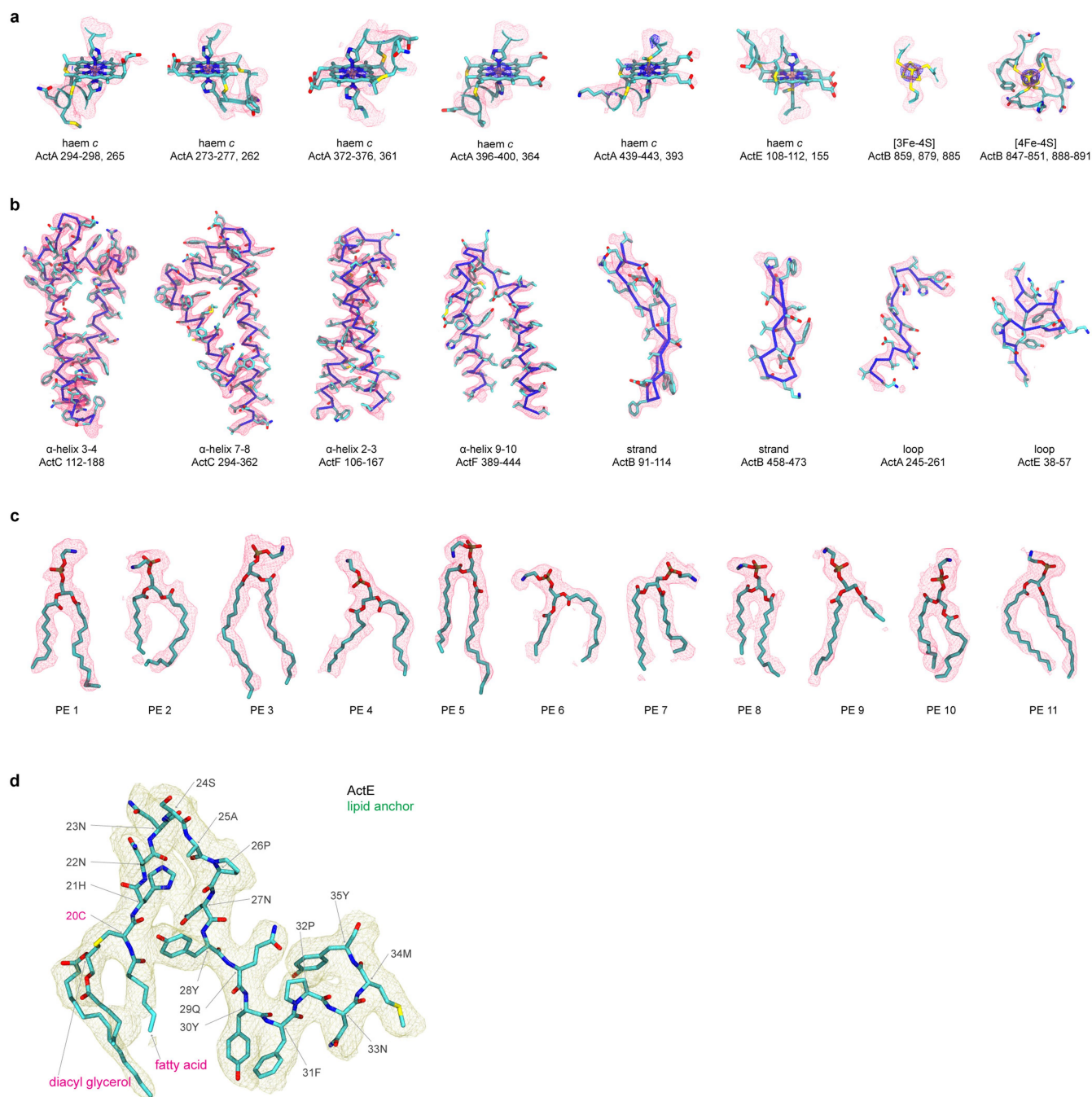
Extended Data Fig. 5 | Features observed in the cryo-EM density and the de novo structure of ACIII. a, Surface representations of ACIII, cyt aa_3 and the ACIII–cyt aa_3 supercomplex. The density threshold is the same for ACIII and cyt aa_3 . **b**, Different views of the ACIII density, coloured by subunit. **c**, Two single-span transmembrane peptides of unknown origin and sequence, denoted ActX and ActY, are present in the structure in the vicinity of ActC. These have each been modelled as a polyalanine peptide. **d**, α -helices 2–10 of ActC form two four-helical up-and-down bundles, coloured in two different shades of blue.

α -helices 1 and 10 are coloured grey and unlabelled. **e**, ActB, shown in cartoon form, has contact with ActA, ActC, ActD, ActE and ActF. Surfaces are drawn from residues that are within 4 Å of ActB and coloured according to their chain. **f**, The transmembrane α -helices of ActC and ActF are arranged in a pseudo two-fold rotation symmetry. **g**, Side-by-side comparison of the polysulfide reductase (PDB 2VPZ) and the assembly of ActB and ActC. These two structures are aligned based on PsrB, the domain containing four iron–sulfur clusters.

a**b****c****Extended Data Fig. 6 | The proposed quinone pocket in ActC.**

a, Sequence alignment of the ActC from *F. johnsoniae*, *R. marinus*, and *Chloroflexus aurantiacus*. The transmembrane α -helices are labelled based on the structure of ACIII from *F. johnsoniae*. The black arrows point to conserved polar residues that are within 15 Å of the [3Fe-4S] cluster in

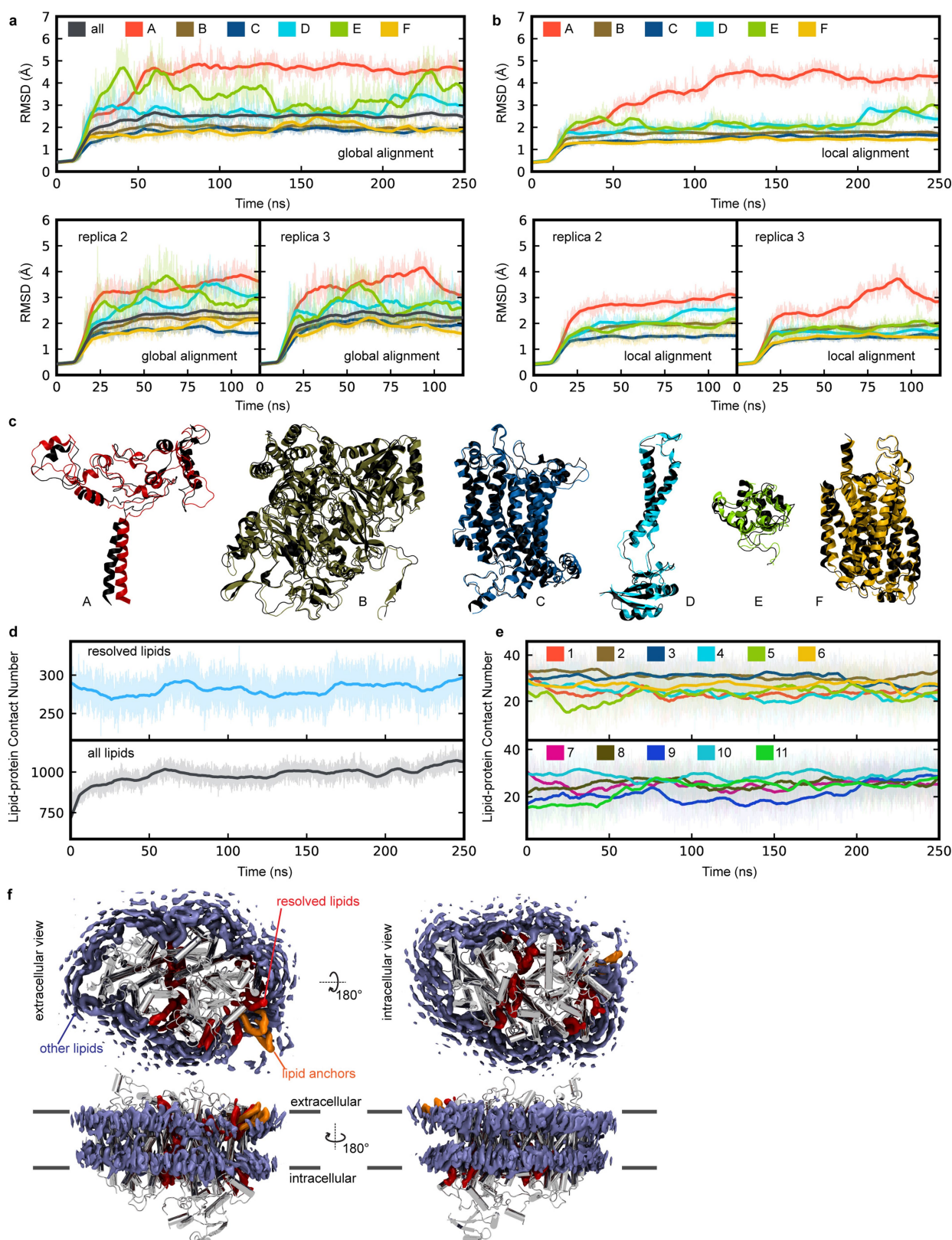
ActB. **b**, Proposed quinone pocket based on the arrangement of conserved polar residues. **c**, Different views of the proposed quinone pocket with a docked menaquinone-1 molecule. Hydrophobic residues near the menaquinone-1 (MK1) head group are also shown. The crevice between α -helix 3 and α -helix 4 is a putative quinone entry pathway.



Extended Data Fig. 7 | Fitting of the ACIII structure to cryo-EM density.

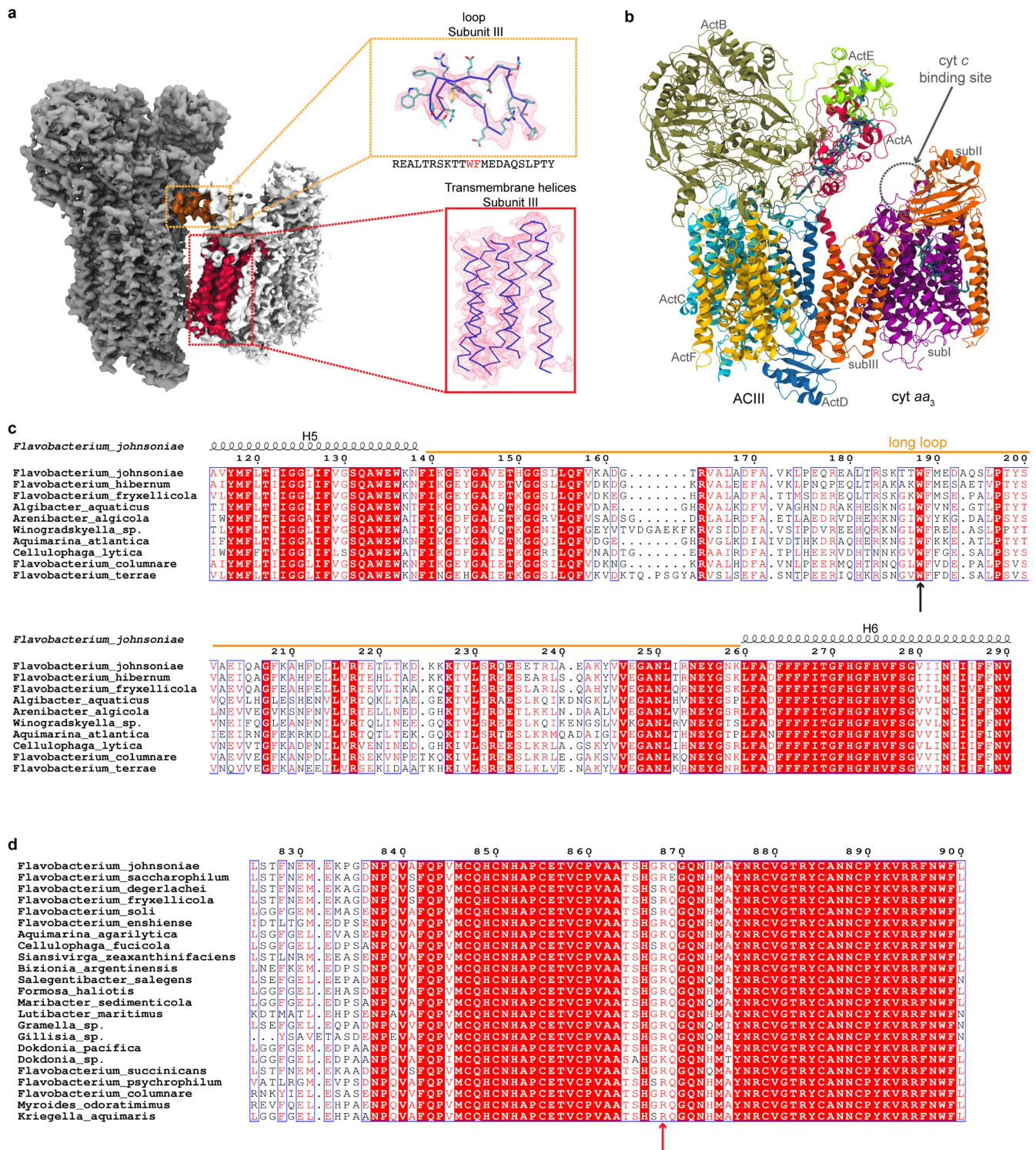
a, Fitting of cofactors into the cryo-EM density. The blue mesh is drawn with a higher density threshold to reveal metal centres. The numberings of nearby amino acid residues, which are shown along with these cofactors, are listed below each cofactor. **b**, Fitting of different secondary structure elements to cryo-EM density. **c**, Eleven identified lipids are modelled as phosphatidylethanolamine molecules. **d**, The triacylated cysteine at the

N terminus of ActE shown along with 15 downstream amino acids. Notably, residue Tyr28 is in contact with the covalent lipid of ActE. Attachment of ActE to the membrane may also be assisted by aromatic residues Tyr30 and Phe31, which appear to be inserted into the lipid bilayer. Throughout the molecular dynamics simulation trajectory, these residues remain buried in the lipid bilayer.



Extended Data Fig. 8 | Protein stability and lipid-protein interaction analysis based on molecular dynamics simulations. **a**, Root-mean-square deviation (r.m.s.d.) of the protein backbone heavy atoms for the entire ACIII complex and each subunit, aligned based on ACIII backbone heavy atoms from three independent molecular dynamics simulations. **b**, Same as **a**, but aligned using the backbone heavy atoms of each subunit. **c**, Superposition of the initial (black) and final (coloured) conformations of each subunit after 250 ns of simulation (aligned using backbone heavy atoms). **d**, The lipid-protein contact number defined by the number of lipid atoms within 4 Å of the protein atoms calculated over the time course

of the simulation. This contact number is either calculated for the eleven lipids resolved by cryo-EM (top) or all membrane lipids (bottom). **e**, The lipid-protein contact number for each of the eleven cryo-EM resolved lipids. **f**, Isosurfaces (50%) of the atom-occupancy map for the lipid anchors (orange), cryo-EM resolved lipids (red) and other membrane lipids (purple), calculated using the last 230 ns of the simulation trajectory. The stronger the lipid-protein interactions, the longer the local residence time, which leads to higher atom-occupancy values. ACIII subunits C, D, and F are shown in silver. For all plots, the raw data are shown as translucent thin lines and the block-averages are shown as dark lines.



Extended Data Fig. 9 | Structural basis for supercomplex formation between the ACIII and the cyt *aa*₃. **a**, Two contact areas between the ACIII and the cyt *aa*₃: the transmembrane portion of subunit III (red) and the loop from subunit III (orange). A homology model of subunit III fits the transmembrane density. The loop is modelled to the cryo-EM density. The sequence of the peptide is also shown. Trp188 and Phe189 are used to register the density to the sequence. **b**, Model of the ACIII–cyt *aa*₃ supercomplex. The cyt *aa*₃ structure from *Rba. sphaeroides* was positioned

based on the transmembrane portion of subunit III. α -helices 1 and 2 of subunit III are omitted to avoid steric clashes with the ACIII structure. **c**, Sequence alignment of subunit III with a long loop (highlighted with the orange bar) between α -helix 5 and α -helix 6 (numbered according to subunit III from *Rba. sphaeroides*). Trp188 (black arrow) is conserved. **d**, Sequence alignment of ActB from organisms with a long loop in subunit III of their cyt *aa*₃ oxidase. Arg868 (red arrow) is largely conserved with occasional substitution to lysine.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	Combined (EMDB-7286) (PDB 6BTM)	ACIII-cyt <i>aa3</i> (EMDB-7447)	ACIII (EMDB-7448)
Data collection and processing			
Magnification	75,000×	75,000×	75,000×
Voltage (kV)	300	300	300
Electron exposure (e ⁻ /Å ²)	61	61	61
Defocus range (μm)	0.8-5.0	0.8-5.0	0.8-5.0
Pixel size (Å)	1.1	1.1	1.1
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	899,405	899,405	899,405
Final particle images (no.)	164,239	81,530	51,547
Map resolution (Å)	3.4*	3.6*	3.6*
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	3.0-4.5	3.0-6.0	3.0-5.0
Refinement			
Initial model used (PDB code)	N/A	N/A	N/A
Model resolution (Å)	3.7 [†]	N/A	N/A
FSC threshold	0.5		
Model resolution range (Å)	3.7	N/A	N/A
Map sharpening <i>B</i> factor (Å ²)	-150.1	-132.4	-129.1
Model composition			
Non-hydrogen atoms	18,935		
Protein residues	2,361		
Ligands	10		
<i>B</i> factors (Å ²)			
Protein	126.4 [‡]		
Ligand	119.7 [‡]		
R.M.S. deviations			
Bond lengths (Å)	0.009		
Bond angles (°)	1.26		
Validation			
MolProbity score	1.36		
Clashscore	1.58		
Poor rotamers (%)	0.15		
Ramachandran plot			
Favored (%)	92.7		
Allowed (%)	7.3		
Disallowed (%)	0.0		

*Determined with cryoSPARC

[†]Determined with Phenix.mtriage[‡]Mean value of the *B* factors determined with Phenix.real_space_refine

TOOLBOX SCIENCE GOES VIRTUAL

Virtual- and augmented-reality tools allow researchers to view and share data as never before. But so far, they remain largely the tools of early adopters.

ILLUSTRATION BY THE PROJECT TWINS



BY DAVID MATTHEWS

As I put on a virtual-reality (VR) headset, the outside world disappears. A cell fills my visual field, and as I crane my neck, I can see it from several angles. I stick my head inside to explore its internal structure. Using hand controllers, I dissect the cell layer by layer, excavating with a flick of the wrist to uncover tiny, specialized structures buried beneath the surface.

Looking at a cell in VR is “as close as you can

get to touching” such a minuscule structure, says Sebastian Konrad, product manager for VR at Arivis, a life-sciences software company in Munich, Germany, that developed this particular VR visualization tool, called InViewR, and who helped to arrange my demonstration of it.

VR isn’t new, but interest in the technology has boomed since 2016, when gamers and a handful of scientists introduced several high-quality, relatively inexpensive commercial headsets to the public. A similar surge has emerged in augmented reality (AR), a related

technology that uses a see-through visor or smartphone screen to layer objects on top of real surroundings.

Some scientists see VR and AR as more intuitive to use than conventional flat screens for viewing complex 3D structures. Others have sought cheap, smartphone-based headsets, which use a smartphone screen as the goggles, to increase public understanding of their work. Their numbers are relatively small: VR and AR remain niche tools for scientific research. Yet some researchers say that the ►

► technology has provided new insights.

Adam Lacy-Hulbert is a principal investigator at the Benaroya Research Institute in Seattle, Washington. He is particularly interested in lysosomes — structures that help to clean up the insides of cells. But he was perplexed by some of the 2D images he was getting using conventional microscopy. “It looked as if part of the lysosomes of the cell had moved into the nucleus, which didn’t really make sense to us.”

But ConfocalVR, a tool developed at Benaroya that uses VR to visualize images from confocal microscopes, made what was really happening “jump out within seconds”, Lacy-Hulbert says. The nucleus was actually deforming and moving around the lysosomes.

Wilian Cortopassi, a postdoctoral researcher at the University of California, San Francisco, has also gained scientific insights from VR explorations. ChimeraX is a molecular-visualization tool for proteins and other structures, which added support for VR headsets in November 2016. ‘Walking’ in virtual space through a network of hydrogen bonds helped Cortopassi to understand how certain mutants of a protein could stymie drugs that target it. A computer monitor is “so messy when you turn on a lot of atoms for visualization”, Cortopassi says. But in VR, “you can just walk through the hydrogens at different angles and distances, and quickly detect some important interactions”.

GOOGLE-EYED

Although inexpensive options are available, most visualization tools work only with the priciest headsets — such as Facebook’s Oculus Rift, and the Vive from Taiwanese electronics company HTC — because they can track the user’s head and handheld-controller movements in 3D space. Researchers and gamers have their preferences, but the differences between Oculus Rift and Vive are small. “I don’t think there’s a clear winner at this point,” says Tom Ferrin, one of the developers of ChimeraX, whose lab at the University of California, San Francisco, specializes in molecular-visualization tools.

That said, not every tool is compatible with all headsets. InViewR works only with Oculus Rift, whereas ChimeraX and ConfocalVR work with both. Oculus Rift and Vive both run using the Windows operating system, although Vive is also compatible with MacOS X.

VR is computationally intensive, both because each eye must see a different image to produce a 3D effect, and because those images must refresh rapidly. In some cases, a new graphics card will add sufficient computing power, “but in general you’re probably going to buy a new computer”, says Tom Skillman, director of informatics and research technology at Benaroya and one of the creators of ConfocalVR. Oculus Rift suggests using VR-compatible computers ranging from US\$850 to nearly \$3,100; it recommends at

least 8 gigabytes of memory and a high-end graphics card.

The VR software itself can also be expensive. Although ConfocalVR and ChimeraX are free for non-profit entities, that is not true for commercial firms. ConfocalVR declined to share pricing information, but ChimeraX can cost up to \$20,000, depending on the number of users.

For researchers who like to work as a team, the developers of ConfocalVR added in April the option for up to four users to simultaneously view, point to and grab structures in the same VR space. This could mean that scientists do not have to meet face to face to work together, says Skillman, which would potentially reduce travel costs. The developers of both ChimeraX and InViewR are looking to add similar collaborative features in the future.

AUGMENTING REALITY

Compared with VR, visualization software for AR headsets is less advanced. Mark Hoffman, chief research information officer at Children’s Mercy Kansas City, a hospital in Missouri, has experimented with viewing proteins and computed tomography (CT) scans using Microsoft’s HoloLens — a kind of visor with a built-in computer that projects 3D objects over the real world.

He says that AR is more user-friendly than VR because users can see their surroundings and so are less prone to disorientation. Hoffman actually experiences motion sickness in VR — and this is not an uncommon complaint. “In all my work with the HoloLens, I’ve never been uncomfortable,” he says.

The downside is that, whereas a VR headset envelops your entire field of view, the HoloLens projects objects only onto a relatively narrow rectangle in the centre of your vision. “It’s part of the trade-off,” Hoffman says. AR is not completely immersive, but it is “an enabler to comprehension”, he says. “There may be things you can miss on a flat screen” that become clearer in AR — protein–protein interactions, for instance.

Surgeons at Children’s Mercy are exploring the use of AR to view CT scans of patients’ hearts before an operation, he says. Hoffman uses a step-by-step approach to make such data viewable using the HoloLens. The surgeon can explore the tissue by projecting it onto a fixed point in space — say, in the middle of the room. But if they turn their head, the image disappears and they see only what is actually there. “They walk into the ventricle or the atrium of the heart, and maybe they’ll see that, for a particular child, the entry point of a blood vessel is not where it normally would be.” The HoloLens

costs \$3,000, and must be ordered from Microsoft directly, because it is not available in the shops.

LOW-COST OPTIONS

Cheaper headsets that use smartphones as the screen in a pair of goggles, such as the Samsung Gear VR or Google’s \$15, ultra-simple Cardboard, can help researchers to reach a broader audience.

Juicebox VR, an app designed for these simple devices, visualizes the connectivity of the human genome as a Mars-like landscape scarred with a colossal wall, says Erez Aiden, a geneticist at Baylor College of Medicine in Houston, Texas, whose lab developed the tool. The features of the landscape represent the topography of condensed DNA in animal cells, and the ridge represents intersections between different parts of the genome. “When people interact with this, they really get a sense of what the data look like,” he says.

Biologists have also adopted Augment, an app normally used to illustrate how furniture might look in a room, to allow colleagues, students and members of the public to inspect 3D models of proteins through their smartphone screens.

For researchers interested in creating their own visualization tools, Unity — software designed by Unity Technologies in San Francisco for building games — is one of the most commonly used development environments. It runs on relatively modest hardware, says Muhammad Saad Shamim, who used it to help to develop Juicebox VR on a Mac Pro. For the HoloLens, users needn’t be advanced developers to import 3D objects, Hoffman says. But they should be comfortable with Unity, as well as Microsoft’s Visual Studio programming environment. Other options include Unreal Engine, from Epic Games in Cary, North Carolina, which is free for academic users, and OpenGL, a no-cost 3D-graphics tool used in game development, computer-aided design and flight simulators. Ferrin, who used OpenGL to create ChimeraX, says OpenGL requires more initial work than Unity or Unreal because developers need to handle more programming details directly, but the pay-off is fewer constraints on functionality.

Despite the broad proliferation of VR and AR tools in consumer culture, only a small minority of labs currently uses the technology, and it remains to be seen how many others will follow suit. Yet many advocates predict that VR and AR could become standard lab tools over the next five years or so. The technology feeds information to our brains in three dimensions, the way “a million years of evolution” intended, says Skillman. It requires an enormous amount of intellectual work to construct a 3D mental model from a 2D screen, he says. “All that work goes away when you put on the goggles.” ■

David Matthews is a freelance writer based in Berlin, Germany.

CAREERS

ESCAPE US political drama drives researcher to move to Canada **p.131**

BLOG Personal stories and careers counselling <https://blogs.nature.com/naturejobs>

NATUREJOBS For the latest career listings and advice www.naturejobs.com



JOHN HOLCROFT/GETTY

ROBBIE HABLE Network for health

PhD student in engineering at the University of Kansas in Lawrence

I was hospitalized for depression in 2017 — and there I learnt the importance of having a support network. It makes your struggles a little bit easier if you have a community. I've reached out to people on campus, but I also found a community thanks to the Cheeky Scientist Association (CSA), a group based in Liberty Lake, Washington, that was created by careers consultant Isaiah Hankel to provide advice and support to researchers worldwide. The CSA posts a lot of success stories, and reminds its members of the value of a PhD. It's been a big help and a huge source of comfort.

I see a therapist weekly. When I walk in, I'm always in a great mood. My therapist validates my emotions and reminds me that I'm mostly struggling against a flawed system, not with a personality flaw. Their continuous encouragement has helped me to focus on finishing my dissertation and keeping my head in the game.

Graduate students are suffering, and they need help. We have fantastic mental-health services on this campus, but a lot of students are hesitant to use them. Some are worried about costs, but they might be surprised. My weekly sessions on campus, for example, are 100% covered by student insurance. Likewise, some students might not want to raise concerns about their adviser or their department out of fear of retaliation. I've been warned not to bite the hand that feeds me.

As president of the campus Graduate Engineering Association, I'm trying to create a sense of community and encourage people to get out of the lab. We threw the first graduate engineering formal in April 2017 at an upscale hotel on campus. There was a DJ, a professional photographer and even a red carpet. We got a huge response. We also have professional-development events. People from industry come here and help us to go over our CVs. We're going through this together.

Like a lot of other students, I struggle with work–life balance. I've faced some criticism for devoting time to my leadership role with the Graduate Engineering Association. But, thanks to my support networks, I have the confidence to be more assertive about my choices. ►

SUPPORT

Why mental health matters

Nature talks to five researchers about the stresses of a hyper-competitive environment, and what needs to change.

BY CHRIS WOOLSTON

More than 150 scientists contacted *Nature* with their personal stories following coverage of an international survey showing evidence of a mental-health crisis in graduate education (T. M. Evans *et al.* *Nature Biotechnol.* **36**, 282–284; 2018). To kick off a series on mental health in academia, we talked to five people on the front lines of science

who were willing to share their insights and discuss how changes to the culture might help.

Next week, we will profile four scientists who have experienced severe depression and its career consequences.

► **NATURE.COM**
Read more stories on mental health in science at
go.nature.com/2hzy8ao

And the week after that, we'll examine health in labs, and ask what kinds of lessons can be learnt from other sectors.

VINCE BUTITTA

Praise for papers

PhD student in limnology at the University of Wisconsin–Madison

I know where my anxiety comes from. Last year, I had a paper come out (V. L. Butitta *et al.* *Ecosphere* 8, e01941; 2017). It was well received and got a lot of attention on Twitter. It was the first time I felt like I was actually doing science, not just playing a part. But then, everything died down. Sometimes I go online to get a figure from my paper, and see that there aren't any new citations. I feel like I'm shouting into the void.

I still struggle with that particular brand of anxiety, but I'm doing what I can to help other students who might feel the same way. When I see a paper that I find interesting, I make sure to send the author an e-mail or message them on Twitter. I say: "I just read your paper — it helped me with some concepts. I look forward to seeing your future work." It lets people know that they have worth. That sort of support doesn't have to come from superiors.

Those messages might help others, but they're also great for me. I connect with other researchers, and when I'm at a conference, someone might recognize my name tag because we've interacted on Twitter.

My paper still doesn't have a lot of citations, but I was invited by a session organizer to speak at an annual meeting of the International Association for Landscape Ecology last month in Chicago, Illinois, because she saw my paper on Twitter. Knowing that someone thought the paper was good enough for a conference gave me a greater sense of satisfaction than did publishing it in the first place.

MATTIAS BJÖRNMALM

Change the culture

Research fellow in materials science, Imperial College London

I'm passionate about protecting and supporting the mental health of early-career researchers. I received my PhD just two years ago, and many people in my life are graduate students or are working with students. I have a personal, emotional connection to their struggles. There's an enormous waste of talent and resources that we're not addressing.

The research culture lies at the core of many scientists' mental-health issues. The environment is hyper-competitive, and the path for success is almost impossibly narrow. That's a scenario that breeds anxiety and depression. People want to produce as much research and as many papers as possible. Anything that takes

away from that can make life more difficult. It is a situation where everyone is pursuing a goal that's almost impossible to reach — the next grant, the next fellowship, the next position.

I'm part of the policy working group for an international professional network called the Marie Curie Alumni Association. I'd like the working group to have a new mission: aligning the incentives and rewards of science with the type of work and productivity that we really want to see. We need to better reward non-traditional outcomes, such as data sets, research methods and code. And we need to better appreciate activities outside of the lab,



such as public engagement, education and outreach. That's the way towards achieving substantial and lasting change.

We also need to encourage students to pursue career prospects both in and out of academia. It's amazing to me how prevalent the belief is that the right path forward is the tenure-track position. People talk about alternative career paths, but too often with the connotation that it's for people who didn't make the cut.

As a scientist, I'm also interested in the evidence. We need to do more to map and monitor the situation. The few studies that have addressed mental-health issues among graduate students had alarming results, but the message isn't getting out. There are still schools that believe they don't have a problem. But anyone who works with graduate students on a day-to-day basis knows that mental-health issues are very prevalent.

For me, and for a lot of people I work with, the whole point of science is to make the world better in some sense. I'm trying to develop new materials as a scientist, but I'm also trying to understand our research culture and how we can improve it. I think it starts with leadership style. If you can create a local environment in your research group or your department that supports talking about these issues and

working on ways to improve them, you can have a big and immediate impact.

We want people to do good research, and we need them to be healthy.

FRANZISKA FRANK

Real-world results

PhD student in ecology and environmental sciences, Umeå University, Sweden

Sometimes I question my worth to society, and this doubt has added to my feelings of depression. Everyone is publishing and publishing because that's where the money in science comes from. But if everyone is publishing and nobody is reading, are we making a contribution? Are we really doing anything important?

There's an excellent service at my university that offers therapy and counselling. You don't have to wait long for an appointment, and they are very familiar with the worries of academics. I've met doctors in the past who didn't seem very interested in the stress that I was feeling. But now, there seems to be more awareness. If we want to talk about depression and mental health, we must acknowledge the progress that has already been made.

Before I started my PhD programme, I did some science communication and education, and that really gave me a sense of satisfaction and validation. We'd take children and their parents to a mobile lab to learn about the North Sea. That's something everyone can relate to. They get really interested in science, and it's not the type of science that comes from a journal.

What do you get from a journal? You submit an article, then you get rejected multiple times. Eventually it's accepted, and you move on to the next thing. Unless you get published in a very prestigious journal or get a lot of citations, it can feel like a downer, even though you accomplished what you were supposed to accomplish.

I would encourage other students to think about what they really want from a PhD. Sort things out for yourself. Talk to people who are important in your personal and professional life, and don't forget to work out. And try to have a life outside the lab.

RACHEL PIPER

Train universities

Policy manager, Student Minds, Oxford, UK

Our charity works with about 120 universities across the United Kingdom. We equip students to cope with graduate school whether or not they've been diagnosed with a mental-health

issue. We have training programmes with university staff members. We want staff members to be able to listen, but they shouldn't be the only source of support for every student.

Our main goal is to make sure that every university has a strategic response to mental health. We support the recommendations of the #stepchange framework, which was launched by Universities UK in 2017 to help improve the mental health of students and faculty members in higher education. Universities must look at their needs and have a specific plan of action to make sure everyone has access to support and treatment.

When the mental-health charity Student Minds started in 2009, many universities denied that they had a mental-health problem on their campus. But the conversation has changed. Now, universities say, 'We know we need to do something, but's what the right thing to do?'

Students also have to look after each other. It's common for people to tell their peers about their troubles but no one else. A 2014 UK study by the Equality Challenge Unit found that 75% of students with mental-health challenges disclosed the issue

"Universities must look at their needs and have a plan of action to ensure everyone has access to support."

to peers (see go.nature.com/2qvhd8k). But according to the Health Education Statistics Agency, only about 3% of all students in the 2016–17 academic year formally reported a mental-health issue to their university. As the discussion continues, hopefully more students will feel comfortable reaching out to supervisors and administrators.

People don't recognize that students have a different experience from other young people. When it comes to National Health Service funding, Student Minds is one of a few groups trying to get more student-health-care models. There's a misconception that students are privileged and don't need extra support. I had my own mental-health concerns as a student, and while I'm much better now, I know how mental health can affect everything. Once you see it, you can't unsee it.

University should be a place where someone can thrive regardless of anxiety or depression. If you have the right support, you can have a diagnosis and still do well. If that support isn't there, you can have no diagnosis and still be stressed. Staff should see university as an opportunity to support people and set them up for their future. If you can help them at university, you're setting them up for a win. ■

INTERVIEWS BY CHRIS WOOLSTON

Interviews have been edited for clarity and length.

TURNING POINT

Political expatriate

Theoretical chemist Alán Aspuru-Guzik was among many US citizens who talked of moving to Canada after the November 2016 election of Donald Trump as US president. Now, Aspuru-Guzik has made good on his declaration, and will begin a new post in July. He explains how the US political climate prompted him to leave his tenured post at Harvard University in Cambridge, Massachusetts, after nearly 20 years in the country.



Why are you leaving the United States?

The nation is at a crossroads. Is it going to continue as a civil society in which politicians and people from different sides respect each other? Or is it going to become a country that has lost political decency and dialogue? Why not use my skills in a country where I don't have to worry about the next national drama, and can concentrate on my science and be with people who share my values?

What will you be doing?

I've accepted a post as a Canada 150 Research Chair in theoretical and quantum chemistry at the University of Toronto, worth Can\$1 million (US\$780,000) a year for 7 years. I'll also be a faculty member at the Vector Institute, which is the new artificial-intelligence research institute in Toronto.

What is the Canada 150 programme?

The Canadian government announced last year that it would invest Can\$117.6 million to enhance the country's "reputation as a global centre for science, research and innovation excellence, in celebration of Canada's 150th anniversary". Canadian institutions get a one-off lump sum to attract top-tier researchers.

What disturbs you most about the US political environment?

We don't have a very civilized way of passing budgets, so even though spending for science was increased, it's tied up with military increases. We have to try to solve climate-change problems. But the United States just left the Paris agreement. I am a dual US–Mexican citizen. I have been here for 20 years, and it doesn't look like it's getting better. Even when the Democrats were in power, the same political war was being waged between the parties. This is the way democracies end — not by coups anymore.

What was it like to work in the United States as a dual national?

I've been lucky to be in some of the most inclusive places in the United States. I lived

roughly half of my time in California and half of it in Massachusetts. I have a PhD; I helped to launch start-up companies; I'm a professor at Harvard and I've published a lot. I'm one of the very privileged in the United States. But how about others who are not? Why should I not worry about them?

Are there drawbacks to vacating your position and leaving collaborators?

I'm leaving a favourable ecosystem. But there are many other great places. Toronto is one. It's one of the most diverse cities in the world, and Canada is leading the world in artificial intelligence and quantum computing. I plan to continue my collaborations at Harvard and the Massachusetts Institute of Technology in Cambridge with key collaborators, and I'll continue to expand them in Canada.

What are your thoughts about leaving the United States and Harvard in general?

All moves are bittersweet. I'm not leaving in any way or form because of Harvard. I'm thankful for them as a platform for my career — they were extremely supportive. Some people believe that one should spend forever in a single place. I think that shouldn't be the case. Sometimes we should do this more often. So I also think it's great that somebody else will take my position at Harvard and that there will be new activity.

What do you see as the main cultural differences between the United States and Canada?

In Canada, people on the street emphasize how welcome you are. And even though you have disagreements, you can still respect your opponents. ■

INTERVIEW BY BRIAN OWENS

This interview was edited for clarity and length.

PIANO CONCERTO FOR NEW HANDS

Time for a change.

BY ANDREA KRIZ

My fingers shake with the effort of pincering the CD. One wrong move, I know, and I'll snap it in two. Slowly, jerkily, I lower it into the player, press it into place and close the lid. No sooner do I do so than my arm thumps back to the covers. Sweat pours down my face. Willing that hunk of metal to move again seems more daunting than climbing Olympus Mons.

"Even in that condition, you're signing on for another tour?"

"There's still fighting on Titan, I'm told," I pant. "Don't worry. After some training, these will feel just like my own."

"I'm not talking about your arms, Cygnus."

"I'm not like you, Halla." I force a smile. "Even if my records have been wiped. I'll never be a real human.

Piloting's the only life I have — I'm trapped."

"You're not trapped."

A nurse opens the door. She drops a bunch of white lilies into a vase, taking in the scene out of the corner of her eye. The charred CD player I'm holding in one hand. My other hand groping for the headphones, finally grasping — only for them to slip through my fingers and clatter to the floor. In one fluid motion, she scoops them up and nestles the buds into my ears.

"You're making progress," she says soothingly.

I close my eyes.

"Who were you talking to?"

Whenever a member of our mech squadron was killed, Halla would do this, mouthing her goodbyes. I didn't even know what this beat-up, orange piece of oldtech was back then. I waited until she'd fallen asleep and dug it out of her locker, pulling on the headphones upside down. Maybe I was expecting some of those crazy Earth drugs I'd heard about. I don't know.

"Oh, you want to listen, too?"

I tried to hear it. But I could never understand those tears that rolled down her face. I read and reread the back of the case. *Ravel. Piano Concerto for the Left Hand.* What did we have to do with a pianist who lost his arm centuries ago? Synthetics like me were biologically incapable of emotion, after all. My previous commander had often told me. And, true to his word, I felt nothing when he died. Even when the engineers determined that we



ILLUSTRATION BY JACEY

crashed because of my shortcomings, when they threw me in a hangar to await incineration. I didn't believe in afterlife. But if I did — it'd be that squadron who landed, butterfly-like, on that dead rock of a colony. The commander who stepped down, haloed in light, said her co-pilot had been gravely wounded, that she'd take anybody, anything, instead.

"I'm Halla."

"I'm a failure."

Her thumb ran over the serial number under my eye. Crossed out. Condemned.

"I don't care about any of that. But it must've been so, so hard for you."

I don't understand. I can't understand. Good synth fight for Earth, get human status. Bad synth fight for separatists, get torched. Why do I keep hearing her voice? I'm only going back where I belong. I reach for the table and knock over the vase. The nurse hurries over and rights the flowers. She settles the tablet into my lap.

"That boy in the synthetic ward you asked me about," she whispers before replacing my earbud. "The one wounded in your last bombing run. He passed away."

The door shuts. Her steps echo down the hall. Should I call her back? I wonder. In my chest, an ache I've never felt before. No. No more thinking. With a swipe of my eyes, I

open the re-enlistment contract, lean in for the signature scan. The strings swell.

"Why? You have so

much to live for. So many people ... waiting for you ..."

"That's right. I've had such a beautiful life, Cygnus. But you. I want you ..."

I spasm, my knees digging into my forehead. My fists, my entire body curls upon itself.

"How can I have a beautiful life?" I whimper. *"When I ended so many ..."*

The piano thunders. I see Halla as I did last. Her head in my lap, blood trickling from her mouth. My palms seared against the twisted wreckage above us. Even as every movement skewered shrapnel into my arms, even as they numbed, I kept clawing upwards. Ignoring the thawing ice pouring in around us. Ignoring the fact that even if I forced my way out of the cockpit, I'd meet enemy fire, I couldn't possibly drag her with me.

"You've gotta promise me, Cyg. I'm leaving you all my CDs."

"How can I listen to music?" I sob. *"I killed them. Synth's just like me."*

"Cygnus. You're always looking up at those stars, aren't you? I can't do this numbers business. That's what I'm calling you."

"They must've wanted names ... they must've had people they loved too ..."

"The war's over, Cygnus."

I feel her through the drumbeat.

You're not trapped. Not trapped.

Slowly, painfully, breathing like the physical therapist instructed me to, I manage to unclench my fists. I can almost see my reflection in their smooth, silver palms. They'd paint them to hide the joints, a rep from the bionics company told me, with plastic just like skin, sculpt every scar and callus into its former place. But maybe I don't want it that way.

These are new hands.

Hands that haven't been trained to kill.

These could be a rocketeer's hands. A cargo-ship driver's. A painter's hands ...

"Halla," I whisper.

I set the tablet down on my bedside table, my grip firm. The last notes of the concerto fade into a victory march, a final chord.

"I think I'll learn to play piano instead." ■

Andrea Kriz flies with the vultures in Cambridge, Massachusetts. Her stories have also appeared in *recompose* and *Diary Science Fiction*.

➔ **NATURE.COM**
Follow Futures:
@NatureFutures
go.nature.com/mtoodm